IJCCS (Indonesian Journal of Computing and Cybernetics Systems)
Vol.13, No.2, April 2019, pp. 177~188
ISSN (print): 1978-1520, ISSN (online): 2460-7258
DOI: 10.22146/ijccs.41259

Text Detection In Indonesian Identity Card Based On Maximally Stable Extremal Regions

Angga Maulana Purba[*1], Agus Harjoko[2], Moh Edi Wibowo[3]
[1]Master Program of Computer Science; FMIPA UGM, Yogyakarta
[2,3]Department of Computer Science and Electronics, FMIPA UGM, Yogyakarta
e-mail: [*1]**angga.maulana.purba@mail.ugm.ac.id**, [2]aharjoko@ugm.ac.id, [3]mediw@ugm.ac.id

Abstrak

Citra hasil scan maupun foto E-KTP yang dimiliki instansi biasanya dikumpulkan dan digunakan sebagai alat verifikasi. Informasi pada foto E-KTP ini dapat dibaca menggunakan software OCR yang banyak berkembang saat ini. Variasi dalam pengambilan foto E-KTP menyebabkan munculnya gambar latar belakang maupun objek selain teks yang tidak diinginkan ditambah dengan orientasi yang bervariasi, hal ini akan berpengaruh pada akurasi pengenalan teks oleh software OCR. Segmentasi daerah teks perlu dilakukan namun tidak lagi bisa dilakukan dengan hanya menggunakan projeksi titik hitam sederhana. Pendekatan yang dilakukan penelitian ini adalah melakukan perbaikan orientasi dari garis yang dibuat oleh Progressive Probabilistic Hough Transform, kemudian deteksi daerah teks dengan algoritme Maximally Stable Extremal Regions yang dikombinasikan dengan horizontal RLSA untuk mendapatkan kandidat daerah teks terbaik dari citra E-KTP. Metode Perbaikan orientasi mencapai rata-rata margin error 0.377o (dalam sistem 360o) dan metode deteksi teks memperoleh akurasi 84.49% pada kondisi terbaik.

Kata kunci— *MSER, Hough Transform, Progressive Probabilistic Hough Transform, RLSA, text detection*

Abstract

Most of Indonesian organizations either it is government or non government sometime required their member to provide their identity card (E-KTP) as legal document collection in their database. This collection of image usually being used as manual verification method. These document images acquired by each person with their own device, there are variations of angles they are used to acquire the image. This situation created problems in text recognition by OCR softwares especially in text detection part, orientation and noise will affect their accuracy. These cases making the text detection more complex and cannot be solved by simple vertical projection profile of black pixels. This research proposed a method to improve text detection in identity document by fixing the orientation first, then using MSER regions to form text region. We fix the orientation using the line that made by Progressive Probabilistic Hough Transform. Then we used MSER to obtain all candidate regions and Horizontal RLSA acts as connector between those candidate. The orientation fixing strategy reach average of margin error 0.377^o (in 360^o system) and the text detection method reach 84.49% accuracy in best condition.

Keywords— *MSER, Hough Transform, Progressive Probabilistic Hough Transform, RLSA, text detection*

Received Nov 24th,2018; Revised April 22th, 2019; Accepted April 25th, 2019

1. INTRODUCTION

Information in identity card, in this case Indonesian identity card is important to all legal organization either government or non government related. Information data in E-KTP or Indonesian identity card can be saved in internal organization database for future use, including verification or statistics purpose. There are many cases in which the organization collecting photo or scan of E-KTP from their members for legal document collection. These collections of image files contains text information that can be extracted through OCR process.

There are many OCR softwares have been developed to complete this process. These softwares accuracy could be degraded by noise or any unwanted objects in image, either it is on Foreground or Background of text documents [1]. The quality of input will affect them, therefore clear text on images without any unwanted object will increase accuracy of these OCR software[2].

The identity card usually has characteristic of neat format and fixed location where the text is. This characteristic enable us to detect the text location in simple way like using vertical projection of black pixels, which had been done before in E-KTP recognition research[3] and the researchers from Bangladesh to accomplished this task[4]. They use vertical projection of black pixels because there were enough gaps between text areas in images. This approach indeed can be used when the photo only contains text and had been taken by camera from close up range.

The problem emerged when this photo had been taken by their owner with variation of distance between the camera and object. There was also problem about the orientation of camera which not fully straight. These cases causing the document images contain any unwanted object or noise which make vertical projection approach can't be used.

The orientation of image could affect OCR's accuracy[5], therefore image orientation fixing is also required. This paper proposed to fix the orientation of images first, before detecting the text region with one of the region detection methods available.

This paper proposed using the line that formed by the text region in image and use it as reference to determine the orientation of image. This approach has been used before for determining Devenagari letter orientation (India and Nepal)[6]. They used Standard Hough Transform to detect the line that formed by middle stroke in the letter which is the main characteristic of Devenagari letter.

The Standard Hough Transform that used before in Devenagari research is computionally expensive because of its voting mechanism. Therefore in this paper we proposed using PPHT (Progressive Probabilistic Hough Transform) as our method to detect the line. PPHT has lower complexity compare to the standard version because it is only using random subset of edge pixels[7]. PPHT can also produce the line length information which can be used as line selection.

After fixing the image orientation, the process continues to text detection process. This paper approach is to detect groups of letter regions which can form text regions in images. This paper proposed using MSER (Maximally Stable Extremal Regions) as main method to detect letter regions. This method usually used in surveillance purpose, i.e. plat number in vehicles[8] or street sign[9]. This method is robust enough for images which have text in it. There was some comparative study of region detector, MSER excelled other methods in most case especially region that has homogeny characteristic and clear boundary [10]. These characteristics suits well with E-KTP documents.

This paper trying to contribute in text detection research area especially in identity card documents that has neat format characteristic such as E-KTP. Obtaining good text detection process is required to obtain a clear input for OCR software, and hoping to improve OCR accuracy later.

2. METHODS

2.1 Image Orientation Fixing

This process is required because in previous explanation, there was some study indicated that the orientation of image could affect OCR accuracy[5]. The orientation of image is determined by the assumption that text region in images could form the longest line possible in image, this line will be the reference to image orientation. First assumption is this line could be formed by the text region, if the photo of E-KTP had been taken in close up range, thus not many unwanted object could formed the false line. This base assumption will be tested later. All the process in orientation fixing step will be explained in the following order.

1. Image retrieval, this process is the first step to get original image from the scan, photo, or other sensors with JPEG or PNG file format.
2. Grayscaling, converting three-channel image into one-channel intensity image.
3. Linear Contrast Stretching, this process could be very important to separating text region from any unwanted object in background, hopefully will increase the accuracy later.
4. Canny Edge Detection, producing edge image that consist of binary indicator 1 and 0, this image will be feed to PPHT (Progressive Probabilistic Hough Line Transform*)*.
5. Horizontal RLSA[11], before edge image that produced by canny be feed into PPHT, this paper proposed RLSA as connector between Connected Components and hopefully connecting group of letters into one line.
6. Finding longest line possible with PPHT (Progressive Probabilistic Hough Transform), The implementation of PPHT could produce line length because there was gap checking between pixels. Another advantage in previous explanation is lower complexity and computationally effective compared to standard version. PPHT overall process shown in Figure 2.

   The process start with accepting edge image with binary indicator as input. Then PPHT will collect all non-zero pixels and pick them randomly in iteration. Iteration start with voter checking, if the picked pixel isn't voter to one particular accumulator box, the process continue. This pixel will vote the accumulator (ρ,θ) in each box like shown in Figure 1. These boxes representing the number of votes and information of voter. As the process running, eventually this little boxes will visually make hills and valleys in 3d perspective, every top of the hill representing a line.

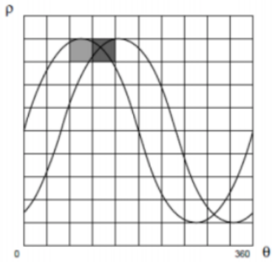


Figure 1 Accumulator (ρ,θ) [12]

Compared to Standard Hough Transform which required all edge pixels to make a valid line. PPHT only used random subset because its verification step to ensure there is no redundancy and thresholding step in voting mechanism. PPHT use thresholding step in case the voter is not enough to considering the line is valid. This process will reduce the iteration process later.

The process continues to check if the line is a good line, the next thing to do is to extract all the voter in corresponding accumulator. These pixels voters will be examined, if there were gaps between those pixels exceeding the threshold then this line will be eliminated otherwise add this line to a "good line". This gap checking will produce beginning point and end point information of a line. This information will be helpful to determine the line length and the orientation of the line.

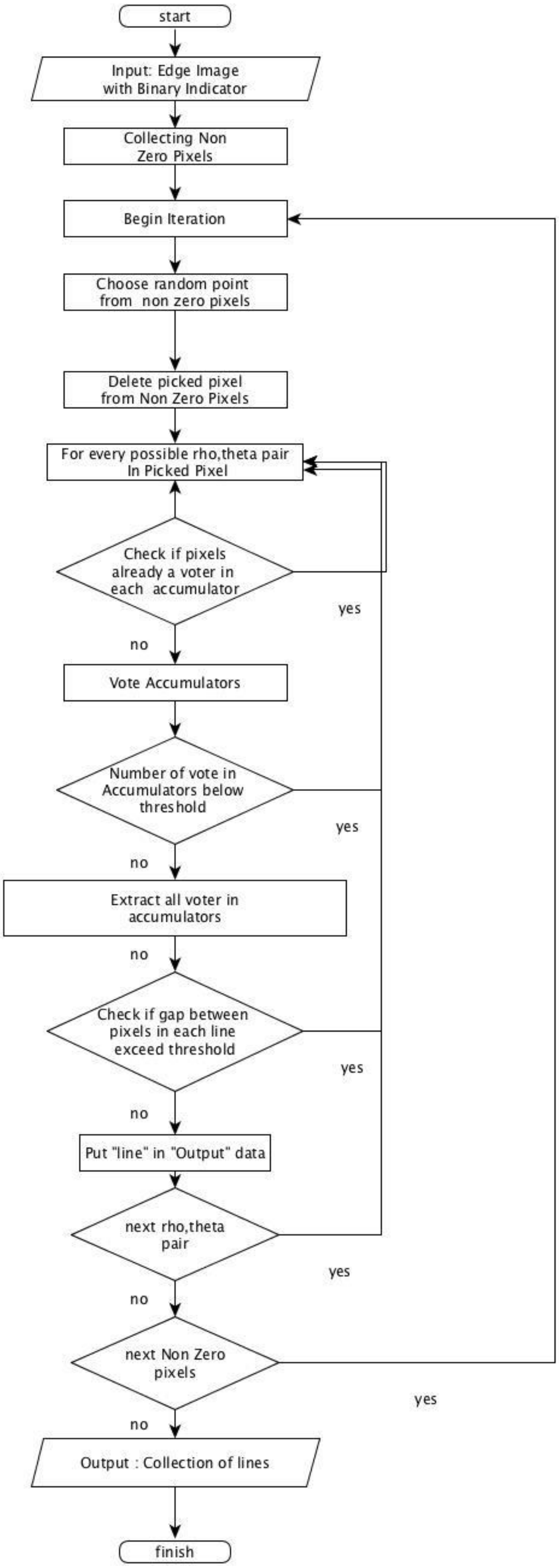

start
Input: Edge Image with Binary Indicator
Collecting Non Zero Pixels
Begin Iteration
Choose random point from non zero pixels
Delete picked pixel from Non Zero Pixels
For every possible rho,theta pair In Picked Pixel
Check if pixels already a voter in each accumulator
yes
no
Vote Accumulators
Number of vote in Accumulators below threshold
yes
no
Extract all voter in accumulators
no
Check if gap between pixels in each line exceed threshold
yes
no
Put "line" in "Output" data
next rho,theta pair
yes
no
next Non Zero pixels
yes
no
Output : Collection of lines
finish

Figure 2 Flowchart Progressive Probabilistic Hough Transform

7. Orientation fixing, after the longest line possible has been found the final process is to determine the angle by calculating the inverse tan of beginning point and end point. Finally, rotate image based on the value of angle calculation.

2.2 Text Detection based on MSER

After fixing the orientation of image, the process continue to detect text regions in E-KTP image. This paper basically proposed similar idea to Enhanced MSER Algorithm [13] with some modifications. This idea basically trying to find groups of neighboring letter region that have small gap. These groups of letter region will be detected by MSER and could form text region later. This paper also proposed RLSA instead of morphology dilation which had been used by Jaswanth[13]. All the proposed process will be explained in the following order.

1. *Grayscaling*, converting three channel image into one channel intensity image.
2. *Linear Contrast Stretching*, this process is the same as previous step and the same image will be used for efficiency.
3. Region detection by *MSER*, overall process shown in Figure 4. The process start with accepting intensity image as input and initializing some variable that will be used later, all the variable needed is the following
   a) Delta, this parameter is used to determine how many iteration the process will take.
   b) MaxArea, this parameter is used to determining maximum size of Connected Components that acceptable.
   c) MinArea, this parameter is used to determining minimum size of Connected Components that acceptable
   d) MaxVariation, this parameter determine the maximum variation that aceptable and considered stable. Every regions hierarchy consists of Extremal Areas that have their own variation value. This variation value calculated by Equation (1), which var_i representing variation of extremals i in every Δ. step.

$$var_i = \frac{Area_{i\text{-}\Delta}\text{-}Area_{i+i}}{Area_i} \qquad (1)$$

Figure 3 shows the illustration of how MSER works through thresholding the images in every step of iteration from white to black, From threshold value 0 causing white image, then eventually black pixels emerged as the process step in iteration. Every delta step will form hierarchy between connected components and considered as parent-child relationship structure.

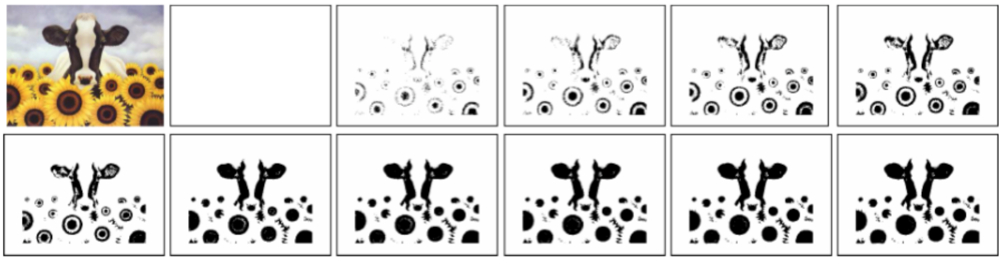


Figure 3 Illustration How MSER Works [14]

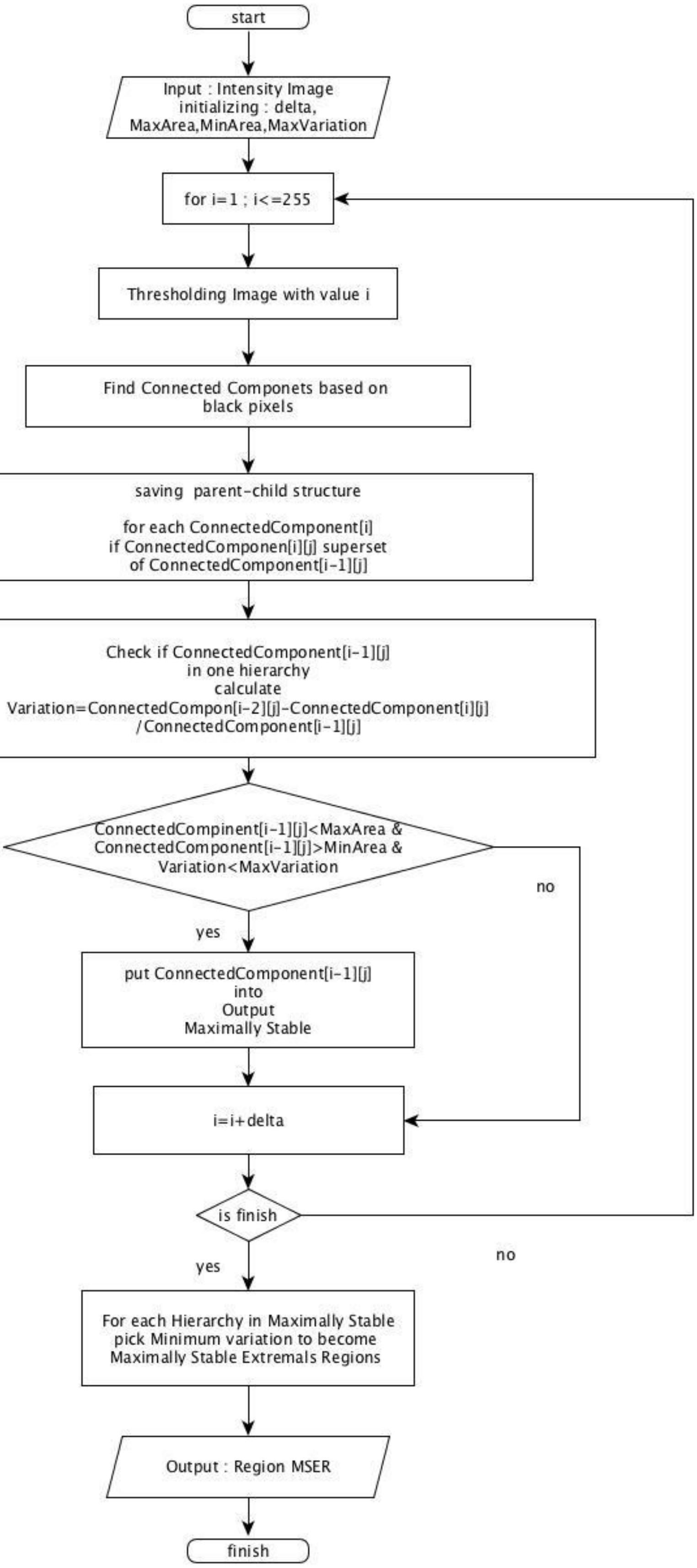
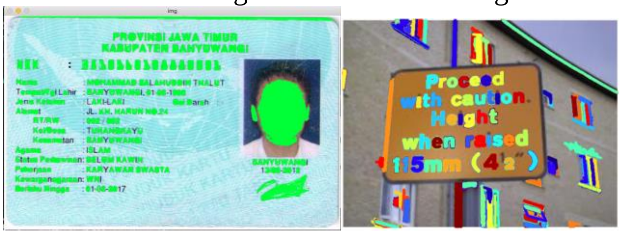

start
Input : Intensity Image
initializing : delta,
MaxArea,MinArea,MaxVariation
for i=1 ; i<=255
Thresholding Image with value i
Find Connected Componets based on
black pixels
saving parent-child structure
for each ConnectedComponent[i]
if ConnectedComponen[i][j] superset
of ConnectedComponent[i-1][j]
Check if ConnectedComponent[i-1][j]
in one hierarchy
calculate
Variation=ConnectedCompon[i-2][j]-ConnectedComponent[i][j]
/ConnectedComponent[i-1][j]
ConnectedCompinent[i-1][j]<MaxArea &
ConnectedComponent[i-1][j]>MinArea &
Variation<MaxVariation
no
yes
put ConnectedComponent[i-1][j]
into
Output
Maximally Stable
i=i+delta
is finish
no
yes
For each Hierarchy in Maximally Stable
pick Minimum variation to become
Maximally Stable Extremals Regions
Output : Region MSER
finish

Figure 4 MSER Process

4. Region Filling, this process also inspired by Jaswanth[13] as shown in Figure 5b and basically trying to filling polygon area that formed by MSER. Every region of MSER is filled by contrast color from the background as shown in Figure 5a



(a) (b)

Figure 5 Region Filling

5. Canny Edge Detection, producing edge image that consists of binary indicator from Region Filled Image.
6. Horizontal RLSA, acts as connector between Connected Components in edge image and hopefully connecting group of letters into one text region.
7. Using bounding box to retrieve all regions the from previous process.
8. Region filter, using minimum and maximum width-height of image and also ratio of width and height as filter to accepting a region as valid.

3. RESULTS AND DISCUSSION

3.1 Linear Contrast Stretching Influence to Delta Finding of MSER and Orientation Fixing

Delta parameter in MSER is used to determine how many iteration will be taken. The larger delta the less number of iteration will be taken, thus affecting complexity in overall process. Finding best delta value is important, to find delta value which not too large or small while maintaining overall performance. Experiment to finding this delta value start with 5 and followed by twice value from previous. This test shows how Linear Contrast Stretching affecting the process of finding best delta. Text regions have intensity close to black and Linear Contrast Stretching trying to separating those regions from the background.

Table 1 Delta Finding without Linear Contrast Stretching

Delta	Accuracy	Precision	Recall
5	0.343100693	0.739380023	0.4025
10	0.52758436	0.819467554	0.615625
20	0.387940235	0.764458465	0.454375
40	0.411544629	0.776992936	0.48125
80	0.410666667	0.773869347	0.48125

Table 2 Delta Finding with Linear Contrast Stretching

Delta	Accuracy	Precision	Recall
5	0.845070423	0.963855422	0.9
10	0.84457478	0.963210702	0.9
20	0.84457478	0.963210702	0.9
40	0.845252052	0.962616822	0.90125
80	0.842413591	0.96187291	0.89875

Table 1 and 2 show how Linear Contrast Stretching affects the experiments of finding best delta. This test shows significant difference between trying to find the delta with and without linear contrast stretching. The best delta with value 40 could reach 0.845252052 in accuracy, 0.962616822 in Precision and 0.90125 in recall. It means the number of iteration only reach 6 at the maximum because the maximum value is 255 divided by 40. This number of iteration is small enough while maintaining the performance.

Table 3 Linear Contrast Stretching Influence to Orientation Fixing

No	Average margin of error (in 360^o system)	Standard Deviation
1	7.427751403^o	17.06576968^o
2	3.817900892^o	17.66674661^o
3	2.60932669^o	13.25430257^o
4	6.592938082^o	21.77099118^o
5	0.377130706^o	0.929846857^o

Experiment result shown in Table 3 indicates how Linear Contrast Stretching also affects the orientation fixing performance. Table 3 shows the average margin of error (in 360 degree system) reduced from first experiment to fifth. Margin of Error is calculated by difference between actual orientation of the image (manual checking) and predicted orientation by proposed algorithm. This experiment gradually stretch the pixels to white or black from the first one to fifth, hopefully could separating text region which closer to black pixel. First experiment reached 7.427751403^o, then in second attempt reduced to 3.817900892^o, in the third 2.60932669^o, then there were anomaly in fourth experiment but finally reached full potential in fifth attempt. This test indicates that separating text regions from the background with Linear Contrast Stretching will also affect orientation fixing performance.

3.2 Orientation Fixing and Text Detection Performances against Image Resolution

The proposed Orientation fixing and Text Detection algorithm also being tested to any kind of image resolution, Low, Medium, and High. Nowadays Smartphones or other gadgets usually have high resolution in their camera, therefore resizing is needed to lowering their quality[15].

Table 4 Text Detection Performance against Image Resolutions

Resolution	Accuracy	Precision	Recall
Low	0.808685446	0.962290503	0.86125
Medium	0.845070423	0.963855422	0.9
High	0.789906103	0.961428571	0.84125

Table 5 Orientation Fixing against Image Resolutions

Resolution	Average Margin of Error	Standard Deviation
Low	0.829885384^o	1.778137495^o
Medium	0.808861769^o	1.781678371^o
High	3.285795938^o	16.26624595^o

Table 4 and 5 indicate that high-resolution photos tend to have bad accuracy because their high detail and quality. Noise and any unwanted object would be clearer to see and affecting region detector. The risk to get more false positive region would be higher because of this high-quality photo, therefore resulting in their worst performance comparing medium or low quality photo and had the worst accuracy, precision and recall among them.

3.3 Reference Line Prediction From Text Area

The orientation of image as in previous assumption (Section 2.1) is determined by the longest line possible in image that created by text area in images. There were some cases in implementation that this assumption was correct. But there were also some cases that the reference line was formed by non text area even though producing the correct result, i.e. The reference line was formed by the top and bottom boundary of E-KTP or was formed by the boundary of profile photo area.

This test as shown in Table 6 indicates that this assumption is correct with accuracy between 68% to 78% in experiments against any kind of resolution. The orientation successfully formed by text area if the photo of E-KTP had been taken with close up range, in this case E-KTP image filled almost all the surface of photo thus making text area so dominant compared to other object. In general, High resolution photo created most problem because of high quality detail and making noise clearer than other resolution, resulting worst accuracy among them.

Tabel 6 Reference Line Prediction From Text Area

Resolution	Accuracy
Low	0.76
Medium	0.78
High	0.68

3.4 Proposed Algorithm Performance against Camera Angles

Table 7 and 8 are the results of the experiments to see how stable both algorithm performed against any camera's point of views. The E-KTP photo had been taken in 9 different angles of camera, like shown in Figure 6 with variety of distances. In general, both algorithm performed quite bad when the distance of camera more than 15 cm because too many unwanted objects in images and created many false positive regions. Both algorithms performed quite well when the distance between the camera and object between 10 cm and 15 cm, the proposed text detection process reached 62% until 72% in accuracy. The proposed orientation fixing algorithm also reached 0.36 until 0.23 (in 360^o system) Average Margin of Error (MoE) in close range distance, as the previous explanation Margin of Error is calculated by difference between actual orientation of the image (manual checking) and predicted orientation by proposed algorithm.

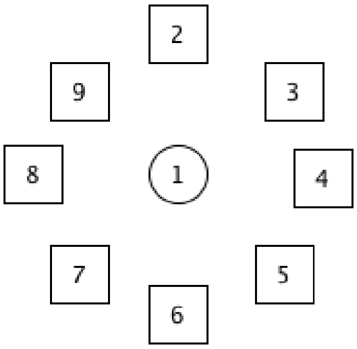


Figure 6 Camera's Point of View (POV) from top

Table 7 Orientation Fixing Performance against Camera's angles

Distance	MoE POV 1	MoE POV 2	MoE POV 3	MoE POV 4	MoE POV 5	MoE POV 6	MoE POV 7	MoE POV 8	MoE POV 9
20 cm	10.3	5.1	45.2	0.3	0.23	15	0.23	0.3	0.23
18 cm	0.24	0.23	0.22	0.15	0.23	0.23	0.21	0.22	0.35
15 cm	0.3	0.23	0.13	0.5	0.23	0.3	0.23	0.23	0.35
10 cm	0.23	0.24	0.23	0.34	0.35	0.23	0.35	0.36	0.35

Table 8 Text Detection Performance against Camera's angles

Distance	Accuracy POV 1	Accuracy POV 2	Accuracy POV 3	Accuracy POV 4	Accuracy POV 5	Accuracy POV 6	Accuracy POV 7	Accuracy POV 8	Accuracy POV 9
20 cm	0.23	0.1	0.25	0.13	0.23	0.15	0.23	0.3	0.23
18 cm	0.34	0.44	0.4	0.28	0.23	0.125	0.43	0.38	0.5
15 cm	0.72	0.62	0.65	0.72	0.75	0.62	0.67	0.73	0.75
10 cm	0.72	0.71	0.25	0.66	0.62	0.68	0.74	0.65	0.62

3.5 Proposed Algorithm Performance in Best Condition

Table 9 indicates the general performance of proposed text detection algorithm in best condition. The photos of E-KTP had been taken from front camera's angle with distance to object between 10 cm and 15 cm. The photos also resized into medium resolution with best setting of Linear Contrast Stretching. It shows that text detection algorithm could reach 84.49% accuracy, with 96.3% precision and 90% recall.

Table 9 Text Detection Confusion Matrix

	Predicted	
Ground Truth	Region	Non Region
Region	1440	160
Non Region	54	50

4. CONCLUSIONS

This study on text detection in E-KTP hopefully contributed to text detection research, especially as an effort to segment text area on image because as shown before in some experiments, it would affect some OCR softwares Performance i.e. tesseract and free-ocr.com. Linear Contrast Stretching had big influence to separating text area from the background thus affecting both proposed algorithm, orientation fixing and text detection based on MSER. The Linear Contrast Stretching process has also significance influence to the effort of finding best delta value for MSER, in order to find ideal delta value which is not too computationally expensive while maintaining the performance.

In orientation problem, Progressive Probabilistic Hough Transform could predict the reference line that was formed by text region quite well, the proposed orientation fixing algorithm could reach average margin of error 0.377^{o} (in 360^{o} system) in best condition. Overall, The Proposed algorithm could perform quite well and reached 84.5% accuracy, 96.3% precision, and 90% recall also in best condition, which is medium resolution photo, front angle camera with distance to object between 10 cm until 15 cm, and best setting of MSER and Linear Contrast Stretching.

This research still has many rooms to improvement. Future works including the weakness of proposed algorithm, i.e. skew photos or perspective fixing and also the risk of false positives because of the lightning problem. The light reflection because of the glossy surface and also uneven light distribution in E-KTP photo will also increase the risk of false positive.

ACKNOWLEDGEMENTS

The author would like to thank Drs. Agus Harjoko, M.Sc, Ph.D and Moh Edi Wibowo, Ph.D. who have guided and provided input in this research, We also thanking LPDP for financial support during study and research.

REFERENCES

[1] A. Farahmand, A. Sarrafzadeh and J. Shanbehzadeh, Document Image Noises and Removal Methods, *International MultiConference of Engineers and Computer Scientists,* Vol I., 2013.

[2] A. El Harraj and N. Raissouni, OCR Accuracy Improvement On Document Images Through A Novel Pre-Processing Approach, *Signal & Image Processing : An International Journal (SIPIJ)*, Vol.6, No.4, 2015.

[3] S. Widodo and Gunawan, "Template Matching pada Citra E-KTP Indonesia", *SNATIKA,* 2015.

[4] R. Akhter, M. Bhuiyandan Uddin., Extraction of Words from the National ID Cards for Automated Recognition, *The International Society for Optical Engineering,* 72-. 10.1117/12.913478, 2011.

[5] N. Jirasuwankul, "Effect of text orientation to OCR error and anti-skew of text using projective transform technique," *IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM),* pp. 856-861., 2011.

[6] T.A. Jundale and R.S. Hegadi, Skew Detection and Correction of Devenagari Script Using Hough Transform, *International Conferenca on Advanced Computing Technologies and Applications*, pp. 305-311., 2015.

[7] A. S. Hassanein , S. Mohammad, M. Sameer, and M. E. Ragab, A Survey on Hough Transform, Theory, Techniques and Applications, *International Journal Of Computer Science*, Vol. 12, Issue 1, 2015.

[8] X. Yang, Y. Zhao, J. Fang, Y. Lu, Y. Zhang and Y. Yuan, "A license plate segmentation algorithm based on MSER and template matching," *12th International Conference on Signal Processing (ICSP)*, Hangzhou, pp. 1195-1199., 2014.

[9] A. Mammeri, A. Boukerche and E. H. Khiari, "MSER-based text detection and communication algorithm for autonomous vehicles", *IEEE Symposium on Computers and Communication* (ISCC), pp. 1218-1223., 2016.

[10] K. Mikolajczyk, T. Tuytelaars , T. Schmid , A. Zisserman, J. Matas, F. Schaffalitzky, T.Kadir, and L. Van Gool, " A Comparison of Affine Region Detectors", *International Journal of Computer Vision*, DOI: 10.1007/s11263-005-3848-x., 2005.

[11] W. Zhu, Q. Chen , C. Wei, Z. Li, A Segmentation Algorithm based on Image Projection for Complex Text Layout, *2nd International Conference on Materials Science, Resource and Environmental Engineering (MSREE)*, 030011-1–030011-8, 2017.

[12] H. Juffry, E. Chandra, and Sofyan, Deteksi Marka Jalan Dan Estimasi Posisi Menggunakan Multiresolution Hough Transform. *Jurnal Teknik Komputer Binus*, 21., 2013.

[13] P. Jaswanth, S. Anusuya, Anil Kumar, and T. Dhikhi , "Enhanced MSER Algorithm for Text Extraction", *International Journal of Computational Intelligence and Informatics*, Vol. 5, No. 4., 2016.

[14] MICC (*Media Integration and Communication Center)*. MSER Presentation lecture, University of Firenze. 2016 [online]. Available : http://www.micc.unifi.it/delbimbo/wp-content/uploads/2011/03/slide_corso/A34%20MSER.pdf . [Accessed: 1-Jan-2018]

[15] E. Christopher and R. Munir, Pengembangan Algoritma Pengubahan Ukuran Citra Berbasiskan Analisis Gradien dengan Pendekatan Polinomial, Konferensi Nasional Informatika., 2013.