Real-Time Detection of Single-Text Character Using the Integration of Extremal Region Filtering and Connected Component Filtering

Karen Niña P. Escosura, Ma. Margarita V. Ortega, Marlen Joyce H. Tuazon, Rovilyn L. Carta, Roselito E. Tolentino

Abstract - Nowadays, among all the contents in images, text information is the most significant value, may it be in a documented text or in a real-world scene. Text detection and recognition in an image is an important task in image analysis. Due to different properties, text detection is a challenging part in an image where textual content are very important. The previous study by Neumann and Matas limits the text detection and recognition in at least three characters. To extend the detection and recognition, especially of single text characters, the proposed solution is to integrate the two methods, Extremal Region Filtering and Connected Component Filtering with Tesseract OCR Engine. Single–text character candidates are filtered by the Connected Component Filtering, wherein the regions extracted and region features from the two stages of Extremal Region Filtering are considered. With structural analysis, connected components with equal value and similar stroke width and stroke orientation are considered single–text character candidates. Character candidates are recognized by the Tesseract OCR engine.

Index Terms -connected component, extremal region, single-text, Tesseract

I. INTRODUCTION

TEXT, in general, carries high-level semantic information. Optical Character Recognition technology allows the conversion of scanned images of printed text into information that can be edited using computer program. The information can be exploited in many content-based image and video applications. Detection of text in an image employs the use of its properties in selection of text. Since texts have different properties, detection of single text characters is challenging due to parting of the foreground and background of an image [1].

The study focuses on image partition in finding text character candidates based on local gradient features and color uniformity of character components and character candidate grouping in detecting text strings based on joint

Revised Version Manuscript Received on March 07, 2016.

Ma. Margarita V. Ortega is currently studying at Polytechnic University of the Philippines Sta. Rosa Campus

Karen Niña P. Escosura is currently studying at Polytechnic University of the Philippines Sta. Rosa Campus

Marlen Joyce H. Tuazon is a graduate of Polytechnic University of the Philippines Sta. Rosa Campus

Rovilyn L. Carta is currently studying at Polytechnic University of the Philippines Sta. Rosa Campus

Roselito E. Tolentino is with Polytechnic University of the Philippines – Santa Rosa Campus and De La Salle University - Dasmarinas as part time Instructor

Structural features of text characters in each text string such as character sizes, distances between two neighboring characters, and character alignment. Using adjacent character grouping method, it calculates the sibling groups of each character candidate as string segments and then merges the intersecting sibling groups into text string. The structure features among sibling characters can be used to determine whether the connected components belong to text characters or unexpected noises. Sibling group of the connected component would search for its left and right sibling set. It will merge the sibling groups into an adjacent character group corresponding to the text string. The character candidate grouping has a capability in detecting text string with at least three characters [3].

The group proposes to innovate a technique that systematically enumerates all possible candidates and confirms each candidate in order to identify which connected pixel regions share the same set of intensity values. The complexity does not depend on the properties of text. It uses geometric constraints to rule out non-text candidates. The group aims to use Connected Component Filtering for pixel region identification in addition to the OCR module. Then the recognized text will be converted into speech using the Microsoft Text-To-Speech Engine. This research is made to help existing text string detection studies and to introduce a method for detecting single text character in real-time image acquisition. The output of this study can be used as assistance to detecting textual characters in some of the following scenes: textual content present in documents, process or step present in some establishment and so on. This can also serve as a basis for further development of single text detection and recognition.

II. METHODOLOGY

The implementation and evaluation of the study is assessed using a desktop computer with an Intel Core i5 (3.5GHz) microprocessor with 4GB RAM running at Windows 7 Ultimate 64-bit. The sensor to be used is A4Tech Anti-glare Webcam PK-835G with a resolution of 640x480 pixels. The proponents' starting distance will be at 10inches. The samples are printed in 5"x4" white paper. Fig. 1 shows the concepts involved upon building the design of the study. The first block is the process of acquiring the image by the camera. The image will then be converted to gray image. The binarized image will undergo ER filtering wherein regions are classified as text string and non-text string regions. Non-text string regions will be classified into two classes as single-text character and non-single-text character candidate by the CC filtering. The classified text string region and single-text character will be considered as the detected text. The Tesseract OCR Engine will recognize the detected text which will be delivered to the Text-to-speech Engine. Hereafter, texts and its corresponding speech will be interpreted.

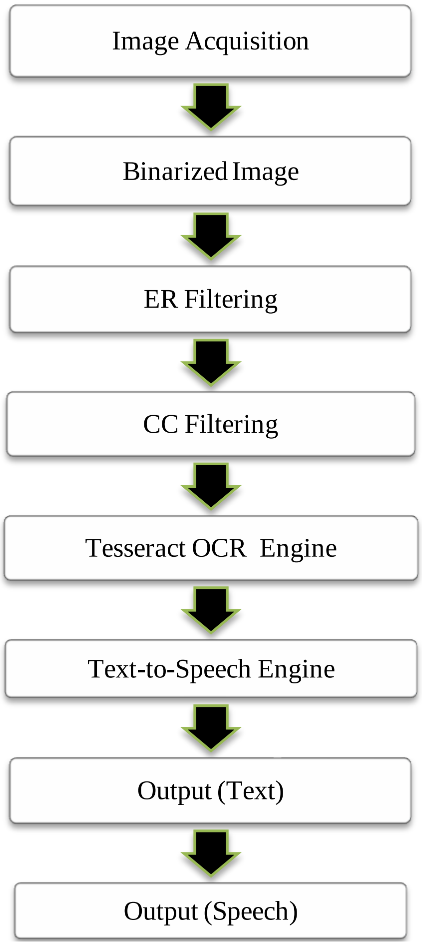


Fig. 1.Conceptual Framework

A. Result of the integration of Extremal Region Filtering and Connected Component Filtering

Performing image acquisition in image processing is always the first step in the workflow sequence. After acquiring the image, it will be converted to a binarized image by three scolding, wherein every pixel in the image is replaced with black and white pixels. The converted image will undergo Extremal Region Filtering.

1. Extremal Region Filtering

Extremal Regions Filtering uses two classifier stages. The first stage make used of Sequential classifier with Real AdaBoost classifier. For classification of Extremal Regions, classifiers need to compute the region descriptors which are the area, bounding box, perimeter, Euler number and horizontal crossing that will serve as its features. The features computed by increasing the threshold will link the connected regions by their inclusion relation [1].

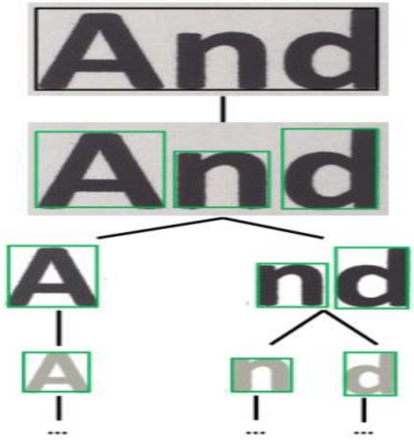


Fig. 2. Representation for inclusion Relation

The classifier will compute the probability if a given region belongs or does not belong to a character candidate according to its region descriptors. Only the ERs which that correspond to local maximum of the probability are selected [3]. The ERs that passed in the first stage classifier will be classified into character. Non-character candidates will proceed to the second stage of ER.

In the second stage classifier, Extremal Region Filtering make used of Support Vector Machines (SVM) and Radial Basis Function (RBF) Kernel. It separates the foreground from the background of the image pixels. It uses the features calculated in the first stage and the complex features, which are the hole area ratio, the convex hull ratio and the number of outer boundary inflexion points [1].

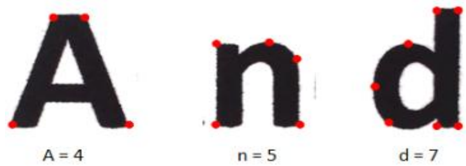


Fig. 3. Numbers of Outer Boundary Inflexion Points for characters

2. Connected Component Filtering

Once regions of non-text string candidates were considered by the ER filtering, they will be fed into the Connected Component Filtering stage. Properties of the regions obtained from the ER filtering will be considered in this stage. The first step is to identify the all the interconnected foreground pixels, starting from the first foreground pixel. Once the first foreground pixel is found, it will be labeled using flood-fill algorithm before going to the next pixel. After all foreground pixels are labeled, they will undergo a connected component extraction.

Starting from the first foreground pixel, connected component extraction is executed by connecting adjacent outer boundary of pixels which have the same value (black or white) and similar stroke width in order to extract connected components or simply mapping of the outer boundary edges of the pixels.

Fig. 4.Edge Mapping of Outer Boundary Adjacent pixels

When the connected components have been extracted, non-character candidate components are eliminated in the stage called connected component verification. This stage rejects connected components that are assigned to a different class (character candidate or non-character candidate) of connected components. After the extraction, the connected pixels will undergo connected component verification, wherein stroke width and stroke orientation is considered. After region extraction and verification, regions are now qualified as single-text character candidates.

3. Tesseract OCR Engine

For the recognition, Tesseract OCR Engine is used. It will analyse the character spacing, text line formation and the height and width of the character to know what character it is [2]. Then, text will be interpreted and converted to speech using Microsoft Speech SDK.

B. Factors affecting the recognition rate

Initially, to evaluate the factors affecting the recognition rate of the system, the system must obtain the recognition rate using the error rate and the total number of characters in one sample.

$$R_{error} = \frac{NumberofCommittedErrors}{NumberofC\ \ haracters} x\ 100\ (1)$$

Recognition rate, which is mostly used for describing the efficiency of a text recognition system, is given by the equation:

$$R_{rec} = \frac{No.ofC haracters\ - No.ofCommittedErrors}{NumberofC\ haracters} x\ 100\ (2)$$

The number of true match is the correct detected text of the given sample. Recognition rate is extracted from each sample given by the proponents. When the recognition of the sample fails in the given interval, the image is considered without any characters present in the image.

III. RESULT AND DISCUSSION

A. Result of the integration of Extremal Region Filtering and Connected Component Filtering

The system uses one hundred (100) samples containing set of words; twenty (20) are single-text characters, ten (10) are two characters but are recognized as single-text characters, five (5) samples are combinations of two characters and single-text character, five (5) are combinations of text string and a single-text character, thirty (30) samples are combination of single-text character and text string in two lines, twenty (20) are combinations of single-text character and text strings in three lines, and ten (10) samples of text strings.

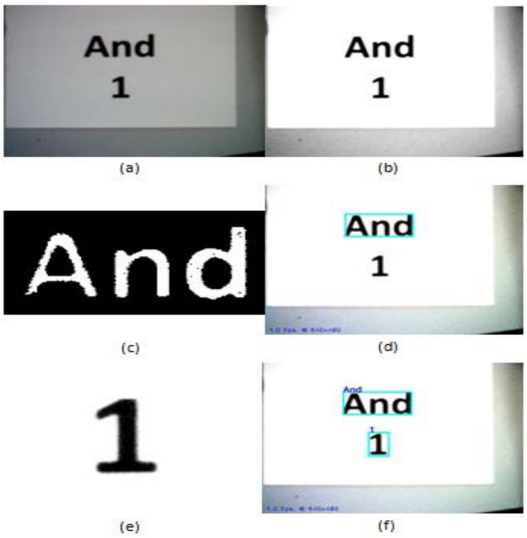


Fig. 5. Data outputs involved in detecting single-text character (a) Source Image (b) Binarized Image (c) Output after passing the First Stage of Extremal Region Filtering (d) Output after passing the Second Stage of Extremal Region Filtering (e) Output after passing the Connected Component Filtering (f) Recognized Text

Fig.5 shows the data outputs of the processes involved in detecting single-text character [1]. Fig. 5 (a) shows the image acquired from one of the samples and binarized in Fig. 5 (b).Figure 5 (c) illustrates the binarized image undergone the first stage of extremal region filtering wherein features are computed from the regions. Fig. 5 (d) demonstrates the second stage of extremal region filtering wherein the regions that will not qualify this stage are considered as non-text string candidates and will undergo connected component filtering. The region recognized by the connected component filtering will be the single-text character candidate, shown in Fig. 5 (e). The classified text string candidate and single-text character candidate will be considered as the detected text. As shown in figure 5 (f), the classified regions, text string candidate and single-text character candidate, resulted from extremal region filtering and connected component filtering are recognized as text characters. Then, these text characters are converted to speech.

B. Factors affecting the recognition rate

The error rate had such a value because of common problems encountered in text detection and recognition in the proposed method, which are character deletion, character substitution and character insertion. The character deletion is caused by character construction, particularly the region width, and character redundancy in one text line. The character substitution and the character insertion are caused by character construction, particularly the similarities of characters and characters with tittle or superscript dot, and indefinite character text line. These problems increase the possibility of a higher error rate thus affecting the accuracy of text detection which results only to an average recognition rate. Regardless of the system having errors, it can still detect and recognize single-text characters using the integration of Extremal Region Filtering and Connected Component Filtering.

Table 1. Recognition Errors

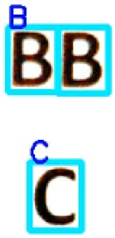
Type of Error	Sample	Remarks
Substitution		No definite text line
Deletion		Character redundancy in one text line
Deletion	 (uppercase i)  (lowercase L)	Region width does not satisfy suitable region width of a character
Deletion/ Substitution		Character construction Tittle (superscript dot)
Deletion and Substitution		Character construction similarity

Table 2
Results of the Study

	Number of Samples	Error Rate	Recognition Rate
Total	100	1215.7609	8784.2391
Average (%)		12.1576	87.8424

Recognition rate for every text detection and recognition application does not have a stable basis for acceptable percentage since there is a wide variety of background, environment and hardware and software used for the system.

Table 1 shows the most frequent errors displayed by the samples based on the data gathered. Individual error rates and recognition rates were computed from each of the 100 samples gathered in ten recognition counts. From the individual rates, the proponents get the average error rate and recognition rate from each recognition count. From this, the total average error rate and recognition rate were obtained. As shown in Table 2, the system's total recognition rate is 87.84% and the error rate is 12.16%

For further improvement, the proponents recommend improving the ability of the system in considering the common factors, such as the hardware and software used in the system, the environment where it was executed and the background where the samples are placed in. Also, the proponents recommend adding error detection and error correction where errors in text recognition could be detected, eliminated, and corrected.

IV. CONCLUSION

Detection of single–text character is made possible by the integration of Extremal Region Filtering and Connected Component Filtering. Character candidates from the image are filtered by the two classifiers of the Extremal Region Filtering. Complex features such as the hole-area ratio, the convex hull ratio and the number of outer boundary inflexion points of the extremal regions are computed to classify the text string character candidates and the non–text string character candidates. Single–text character candidates from the non–text string character candidates are filtered by the Connected Component Filtering, wherein the regions extracted and region features, such as the incrementally computed descriptor features and the additional features, from the two stages of Extremal Region Filtering are considered. Connected components are extracted by connecting the adjacent outer boundary of the pixels. With structural analysis, connected components with equal value and similar stroke width and stroke orientation are considered single–text character candidates. Character candidates are recognized by the Tesseract OCR engine and converted to speech by interfacing the Microsoft Speech SDK in Microsoft Visual Studio.

From the 100 samples, there is a total average of 87.84% recognition rate with 12.16% error rate. The error rate had such a value because of common problems encountered in text detection and recognition in the proposed method, which are character deletion, character substitution and character insertion. The character deletion is caused by character construction, particularly the region width, and character redundancy in one text line. The character substitution and the character insertion are caused by character construction, particularly the similarities of characters and characters with tittle or superscript dot, and indefinite character text line. These problems increase the possibility of a higher error rate thus affecting the accuracy of text detection which results only to an average recognition rate. Regardless of the system having errors, it can still detect and recognize single-text characters using the integration of Extremal Region Filtering and Connected Component Filtering.

V. RECOMMENDATION

For the improvement of the system, the proponents recommend improving the ability of the system in considering the common factors, such as the hardware and software used in the system, the environment where it was executed and the background where the samples are placed in, affecting the detection and recognition of characters. Also, the proponents recommend adding error detection and error correction where errors in text recognition could be detected, eliminated, and corrected. Examples of these are lexicon-based or dictionary-based error correction or based on word construction probability.

ACKNOWLEDGMENT

The proponents would like to extend their deepest and sincerest gratitude to the following for their invaluable support to complete this study:

First, the proponents would like to express their deepest faith and gratitude to the Lord God whose presence may not have seen physically but is always there to send His Holy Spirit for continuous guidance and blessings;

Second, to their dearest families and relatives, who in spite of all their shortcomings are still there to give them financial, moral support, buttress of love and understanding; Third, to Engr. Roselito E. Tolentino, their Thesis adviser, for generously sharing his expertise, for his guidance and continuous support throughout the cause of completing this paper. Without his guidance and help, this paper will not be possible; and Lastly, the proponents would also like to extend their gratitude, to all of their friends and classmates, who stood with them through thick and thin, and comfort them during their distress.

REFERENCES

1. Neumann, L. and Matas, J. (2012). Real-Time Scene Text Localization and Recognition. USA: 25th IEEE Conference on Computer Vision and Pattern Recognition.
2. Smith, R. (2007). An Overview of the Tesseract OCR Engine. IEEE
3. Tian, Y. and C. Yi (2011). Text String Detection from Natural Scenes by Structure-based Partition and Grouping. IEEE,.
4. Eikvil, L. (1993). OCR Optical Character Recognition.
5. Matas, J., Chum, O., Urban, M. and Pajdla T. (2004). Robust Wide Baseline Stereo from Maximally Stable Extremal Regions.
6. Halgaonkar, P.S. (2011). Connected Component Analysis and Change Detection for Images. India: International Journal of Computer Trends and Technology.
7. Mithe, R., Indalkar, S. and Divekar, N. (2013). Optical Character Recognition. International Journal of Recent Technology and Engineering.
8. Neumann, L. and Matas, J. (2011). Text Localization in Real-world Images using Efficiently Pruned Exhaustive Search. Czech Republic: Czech Technical University.
9. Saabni, R. andZwilling, M. (2012). Text Detection and Recognition in Real World Images. International Conference on Frontiers in Handwriting Recognition

AUTHOR PROFILE

Karen Niña P. Escosura was born in Sta. Rosa, Laguna, Philippines on April 9, 1995. She received the degree in Bachelor of Science in Electronics and Communications Engineering from the Polytechnic University of the Philippines Santa Rosa Campus, Laguna in 2016. She was a part of Panasonic Manufacturing Philippines Corporation (PMPC) in City of Santa Rosa, Laguna, Philippines as a team member of the Factory Engineering Department for her On-the-Job training in 2015. Ms. Escosura is currently a member of Institute of Electronics Engineers of the Philippines and Association of Electronics and Communication Engineering Students.

Ma. Margarita V. Ortega was born in Biñan, Laguna, Philippines on October 24, 1995. She received the degree in Bachelor of Science in Electronics and Communications Engineering from the Polytechnic University of the Philippines Santa Rosa Campus, Laguna in 2016. She was a part of ON Semiconductor, Carmona, Cavite, Philippines as a team member of the TPED Department for her On-the-Job training in 2015. Ms. Ortega is currently a member of Institute of Electronics Engineers of the Philippines and Association of Electronics and Communication Engineering Students.

Marlen Joyce H. Tuazon was born in Sta. Rosa, Laguna, Philippines on August 8, 1995. She received the degree in Bachelor of Science in Electronics and Communications Engineering from the Polytechnic University of the Philippines Santa Rosa Campus, Laguna in 2016. She was a part of Amkor Technology Philippines, Inc., City of Santa Rosa, Laguna, Philippines as a team member of the Test Department for her On-the-Job training in 2014 and in NXP Semiconductors, Inc in City of Cabuyao, Laguna, Philippines as a team member of the IT and Communications Department for her On-the-Job training in 2015. Ms. Tuazon was currently a member of Institute of Electronics Engineers of the Philippines and Association of Electronics and Communication Engineering Students.

Rovilyn L. Carta was born in Sta. Rosa, Laguna, Philippines on September 16, 1994. She received the degree in Bachelor of Science in Electronics and Communications Engineering from the Polytechnic University of the Philippines Santa Rosa Campus, Laguna in 2016. She was a part of Panasonic Manufacturing Philippines Corporation (PMPC) in City of Santa Rosa, Laguna, Philippines as a team member of the QAC Department for her On-the-Job training in 2015. Ms. Carta was currently the Secretary of the organization Association of Electronics and Communications Engineering and a member of Institute of Electronics Engineers of the Philippines and Association of Electronics and Communication Engineering Students.

Roselito E. Tolentino is a registered Electronics Engineer and IECEP-Member. He is a graduate of B.S. Electronics and Communication Engineering at Adamson University in 2004 under the scholarship of DOST-SEI. He finished his Master of Science in Electronics Engineering Major in Control System at Mapua Institute of Technology under the scholarship of DOST-ERDT. He currently takes up Doctor of Philosophy in Electronics Engineering at the same Institute. He is currently working as a part time instructor at Polytechnic University of the Philippines Santa Rosa Campus and De La Salle University - Dasmarinas. His research interests are more on Robotics and Machine Vision.