

Text Region Extraction in High Contrasting Image

Seokjun Kang, Daewoong Cha, Youngwoo Kim, and Dong Seog Han

Abstract—In this paper, a text region extraction system with high contrasting text images for self-driving cars is proposed. The maximally stable extremal regions (MSER) method is usually used to extract text regions. Images must be converted to grayscale to process with the MSER method. However, the performance of MSER by using grayscale images has a poor ability of capturing regions of interest in bad conditions such as high-contrast, low-luminance, much light reflection, and so on. An MSER system with a contrast-limited adaptive histogram equalization (CLAHE) instead of conventional MSER is therefore proposed. CLAHE is utilized as a pre-processing method in MSER to detect text regions. The proposed method achieves a precision of 81% and a recall of 82%. However, those for the MSER with grayscale are 63% and 55%, respectively.

Index Terms—CLAHE, MSER, stroke width transform, text region extraction.

I. INTRODUCTION

In self-driving cars and advanced driver assistance systems (ADAS), camera sensors can be used for the detection of vehicles, pedestrians, and traffic signs. Written texts on road sign posts provide lots of information for safety driving. However, recognizing the text information is not easy in bad weather or light conditions causing high-contrast, low-luminance, and much light reflection. To recognize text information, the text region must be extracted. Extracting a text candidate region in an image is a difficult task because the region has a small portion of an entire image. Examples of images that are difficult to detect text information are shown in Fig. 1. So, there is a need to provide solutions to the text detection problem for road signs. Thus, conventional text detection systems need to be studied and improved for road signs.

This paper proposes an enhanced text information detection method for high contrasting text images. In order to detect text information in images, text candidate regions must be extracted first. To detect the text candidate regions, the maximally stable extremal regions (MSER) method [1]-[6] is used. In the MSER process, the grayscale images are used.

In this paper, the contrast-limited adaptive histogram equalization (CLAHE) [7] is used to improve the text detection performance. In addition, the Canny edge detection [8] is also used to guarantee the results obtained with MSER. We merge the MSER and Canny edge results to segment text

Manuscript received February 6, 2017; revised March 20, 2017. This research was supported by The Leading Human Resource Training Program of Regional Neo industry through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and future Planning (grant number: NRF-2016H1D5A1909920).

Seokjun Kang, Daewoong Cha, Youngwoo Kim, and Dong Seog Han are with the School of Electronics Engineering, Kyungpook National University, Daegu, Korea (e-mail: dshan@knu.ac.kr).

regions. Finally, we carry out the stroke width transform (SWT) method [9], [10] to detect the final text region. We compare our proposed system with the conventional MSER system.

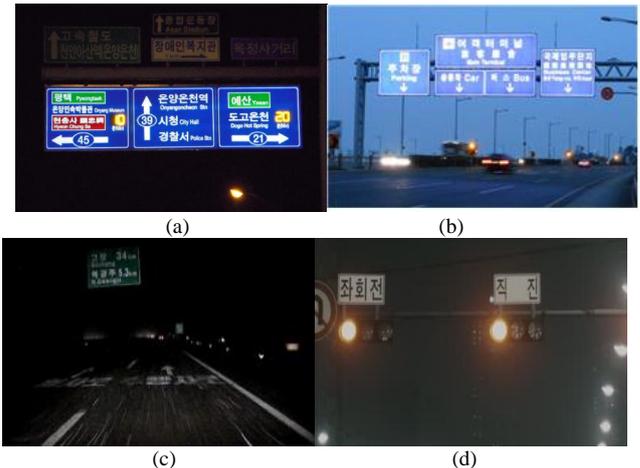


Fig. 1. Poor environment of text detection. (a) high-contrasting image (b) image at night (c) reflection effect (d) blurred image.

The rest of this paper is organized as follows. Section II describes minutely the process of extracting text regions. Section III presents experimental results for detecting text regions. Section IV presents the conclusion.

II. TEXT REGIONS EXTRACTION

In this section, the process of extracting text regions is elaborated. First, the CLAHE approach is explained and then its application is proposed for text region detection. The MSER algorithm is presented and its role in the proposed system is described. The Canny edge detection process description and its application to enhance system performance follows. The reasons for using the Canny edge detection method are also elaborated. Finally, the SWT algorithm that detects the final text region is presented.

A. Procedure from CLAHE to MSER

The conventional MSER utilizes grayscale or histogram equalization for pre-processing. The purpose of histogram equalization is to spread the brightness of pixels in the input images. This is done so that dark and contrasting images can be further improved for better object detection performance. Unfortunately, the method is not good enough to recognize text regions. In order to solve the problem, the CLAHE method is adopted in this paper.

CLAHE limits the contrast of an image using histogram equalization. CLAHE differs from ordinary adaptive histogram equalization in its contrast limiting. This feature can also be applied to global histogram equalization, giving rise to contrast limited histogram equalization (CLHE). In the

case of CLAHE, the contrast limiting procedure has to be applied for each neighborhood from which a transformation function is derived. CLAHE was developed to prevent the over-amplification of noise that adaptive histogram equalization gives rise to.

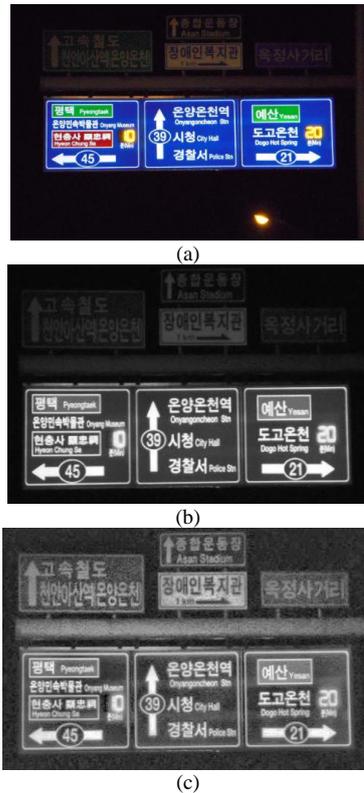


Fig. 2. The result of CLAHE and grayscale. (a) Original image (b) Result of grayscale method (c) Result of CLAHE method.

Two parameters must be considered to limit the contrast of an image using the CLAHE method. One is clip-limit and the other is block-size. The contrast amplification in the vicinity of a given pixel value is given by the slope of the transformation function. This is proportional to the slope of the neighborhood cumulative distribution function and therefore to the value of the histogram at that pixel value. CLAHE limits the amplification by clipping the histogram at a predefined value before computing the cumulative distribution function. The value at which the histogram is clip-limit, depends on the normalization of the histogram and thereby on the size of the neighborhood region. So, the inordinate contrast of the image should be controlled by setting the clip-limit value. Finally, clearer images are realized. In CLAHE, block-size is one of the parameters. If a block-size is decreased, the contrast of an image is decreased too. Those parameters are got by a trial and error based learning procedure. In our proposed system, a clip-limit to 0.02 and a block-size to $16 \times$ was chosen. The results of CLAHE are shown in Fig 2. As shown in Fig. 2, the road milestone are visible and invisible. This is a problem caused by high contrast of an image. To solve this problem, we can confirm that it is appropriate to apply CLAHE by checking the road milestone in the image.

In the proposed system, the MSER algorithm is executed using the result of CLAHE. The MSER algorithm is mainly used for text detection because it shows excellent performance for blob detection in an image. Detecting text

regions by using the MSER method can be summarized as following four steps:

- Step I: thresholding a luminance of an image.
- Step II: extracting the extremal regions or connected components.
- Step III: evaluating whether extremal regions are maximally stable.
- Step IV: obtaining text regions.

CLAHE and MSER are used to detect the primary text candidate region.

B. Canny Edge Detection

In order to extract text contours, Canny edge detection is used in this paper. The Canny edge method has four stages to get an edge map. The first stage, we smooth images by applying the Gaussian filter which decreases the noise effect. In the second stage, we use the Sobel-edge operator to get the intensity gradient of the image. The third stage is applying the non-maximum suppression. The non-maximum suppression algorithm is first compares the edge strength of the current pixel with those of the pixels in the positive and negative gradient directions. Finally, weak edge lines are eliminated by applying hysteresis thresholding. Results of edge detection for Sobel and Canny are shown in Fig 3. Using the edge detection results in Fig. 3, the Canny-edge operator preserves more of the edge information compared to the Sobel edge operator. Therefore, the Canny edge operator is chosen instead of the Sobel edge operator.

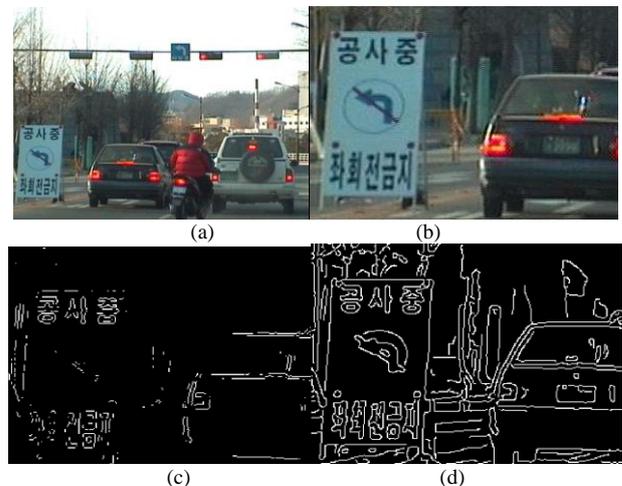


Fig. 3. Results of edge detection. (a) Original image (b) Enlarged image (c) Result of Sobel edge detection (d) Result of Canny edge detection.

C. Detection for Text Candidate Regions

The purpose of our proposed system is to find the text candidate regions as accurately as possible. First, we find the text candidate regions by applying MSER with the CLAHE algorithm result. At the same time, we perform Canny edge detection with grayscale images. Although the MSER algorithm is an excellent method for text detection, there is a problem of blob detection rather than text information. Therefore, in order to detect a more accurate text candidate region in the image, we detect not only the MSER but also the Canny edge, and combine them to obtain a text candidate region from which a lot of blobs are removed. The merged result is changed to the final text region using the SWT algorithm. Our purposed overall system is shown in Fig. 4.

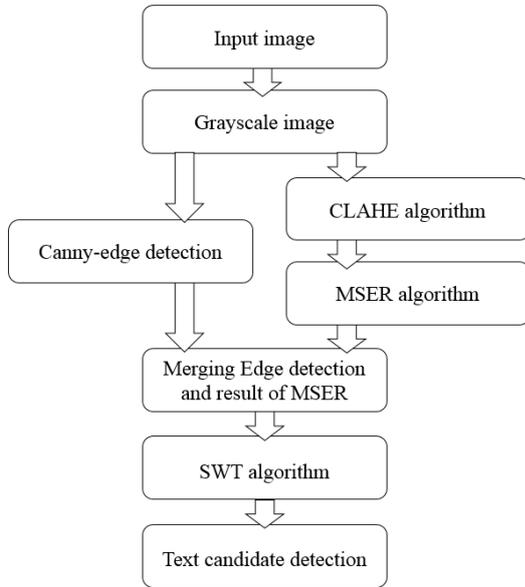


Fig. 4. The proposed overall system of text regions extraction.

D. Extraction of Final Text Regions

The SWT algorithm is used to detect the final text regions as follows. The width of the most likely stroke containing the pixel is calculated for each pixel. To obtain the real text regions, we proceed with the partial differential operation at the edge lines. And then, we obtain the same gradient directions in real text regions because almost texts compose a uniform width and size. Using this information, the stroke width corresponding to the text is detected by utilizing the gradient direction of the edge. Finally, the real text regions are obtained. The SWT process of the proposed system is shown in Fig. 5.

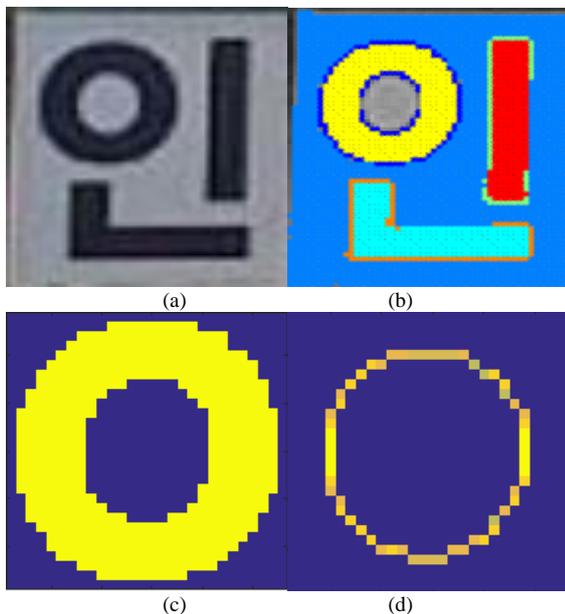


Fig. 5. Process of SWT. (a) Original image (b) MSER detection result (c) Text region image with scaled color (d) Stroke width image.

III. EXPERIMENTAL RESULTS

In order to evaluate the performance of the proposed system, computer simulations have been conducted. We create our own dataset for the experiment. Our dataset is as

follows. It is a set of 120 image datasets, including 60 road milestone images extracted from good illuminance environments and bad environments such as high contrast, respectively. The results of text detection are shown in Fig. 6. The performance of CLAHE and other methods used for pre-processing are compared when applied with the MSER method. The accuracy of the proposed system is presented in Table I. The precision and recall [11] are calculated as in (1):

$$\text{Precision} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalsePositive}} \quad (1)$$

$$\text{Recall} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalseNegative}}$$

Table I shows the results of our proposed algorithm. Recall is a measure of how well objects are detected without missing. On the other hand, precision is a value indicating how many objects to be detected are included among the detection results. Although the recall value is important in a system that detects a vehicle or a pedestrian, the precision value is important due to problems caused by misdetection of the object. However, in the text detection system, it is very important to detect the candidate region of characters in the image, so the value of recall is more meaningful.

TABLE I: RESULT OF THE PROPOSED ALGORITHM

Total Frames	True Positives	False Positives	False Negatives	Precision	Recall
120	89	21	20	0.81	0.82

TABLE II: RESULT OF MSER WITH AND WITHOUT CLAHE

Method	True Positives	False Positives	False Negatives	Precision	Recall
MSER with CLAHE	89	21	20	0.809	0.817
MSER without CLAHE	59	32	29	0.63	0.55

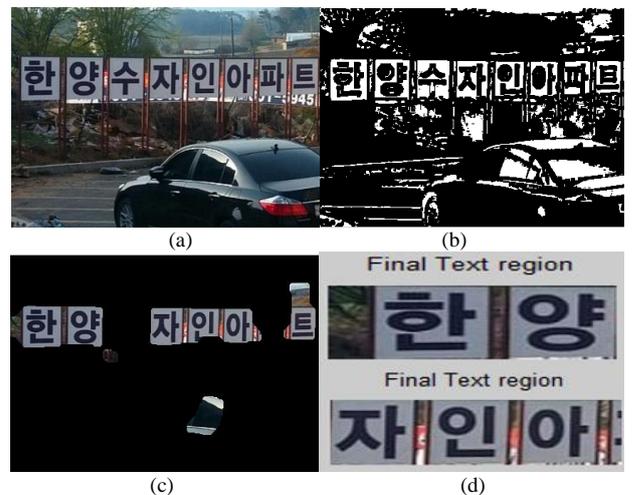


Fig. 6. Text region detection. (a) Original image (b) Merged MSER and Canny edge-detection (c) Text region candidate extraction (d) Final text region detection.

In order to compare the performance of the proposed algorithm, we compared the results of the proposed algorithm

with the conventional algorithm without CLAHE. Detecting candidate character regions using MSER shows good performance, but shows poor performance in bad environments such as high contrast.

Table II shows the performance of the algorithm with and without CLAHE. In Table II, the proposed system realizes a higher value of precision and recall than the conventional system. The high value of recall means that the text regions cannot be found in the conventional text region detection. Eventually, it is easier to detect the character in the whole image by adjusting the contrast of the image by applying CLAHE.

IV. CONCLUSION

In this paper, we proposed a text region detection system. The proposed system applies the CLAHE algorithm to obtain good performance in high contrast images, and merges the result of Canny edge detection and MSER to eliminate meaningless blobs detected. In this paper, a method to detect text regions in high contrast images was proposed and presented. The text information on road milestones is essential information to provide driver safety and comfort in ADAS, such as autonomous driving, however it is not easy to detect. We have recognized that problems such as the difficulty of detecting text information due to poor management as well as high contrast images should be further studied. To this end, we intend to design an integrated text detecting system which utilizes fully convolutional network (FCN).

REFERENCES

[1] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in *Proc. the British Machine Vision Conference (BMVC)*, pp. 384-393, Sep. 2002.

[2] S. Lee, M. S. Cho, K. Jung, and J. H. Kim, "Scene text extraction with edge constraint and text collinearity," in *Proc. 20th International Conference on Pattern Recognition (ICPR)*, pp. 3983-3986, Aug. 2010.

[3] L. Gómez and D. Karatzas, "MSER-based real-time text detection and tracking," in *Proc. 22nd International Conference on Pattern Recognition (ICPR)*, pp. 3110-3115, 2014

[4] M. R. Islam, C. Mondal, M. K. Azam, and M. J. Islam, "Text detection and recognition using enhanced MSER detection and a novel OCR technique," in *Proc. 2016 5th International Conference on Informatics, Electronics and Vision (ICIEV)*, pp. 15-20, May 2016.

[5] X. C. Yin, X. Yin, K. Huang, and H. W. Hao, "Robust text detection in natural scene images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 970-983, May 2014.

[6] J. Greenhalgh and M. Mirmehdi, "Real-time detection and recognition of road traffic signs," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1498-1506, Dec. 2012.

[7] K. Zuiderveld, "Contrast limited adapted histogram equalization," *Graphics Gems IV*, Academic Press, San Diego, 1994.

[8] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 679-698, Nov. 1986.

[9] B. Epshtein, E. Ofek, and Y. Wexler, "Detecting text in natural scenes with stroke width transform," in *Proc. 2010 Computer Vision and Pattern Recognition (CVPR)*, pp. 2963-2970, Jun. 2010.

[10] Y. Li, H. Lu, "Scene text detection via stroke width," in *Proc. 21st International Conference on Pattern Recognition (ICPR)*, pp. 681-684, Nov. 2012.

[11] S. Kim, E. Jang, S. Hyun, and D. S. Han, "Real time object detection based on saliency map," in *Proc. 2016 IEEE International Conference on Consumer Electronics (ICCE)*, pp. 534-535, Jan. 2016.



Seokjun Kang received his B.S. degrees from Kyungpook National University, Daegu, Korea, in 2016. He is currently pursuing his M.S. degree. His main research interests are image processing and machine learning.



Daewoong Cha received his B.S. degrees from Kyungpook National University, Daegu, Korea, in 2016. He is currently pursuing his M.S. degree. His main research interests are signal processing and active sonar.



Youngwoo Kim received his B.S. degrees from Kyungpook National University, Daegu, Korea, in 2016. He is currently pursuing his M.S. degree. His main research interests are signal processing.



Dong Seog Han (SM'07) received his B.S. degree in electronic engineering from Kyungpook National University (KNU), Daegu, Korea, in 1987, and his M.S. and Ph.D. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 1989 and 1993, respectively. From October 1987 to August 1996, he was with Samsung Electronics, Co. Ltd., where he developed the transmission systems for QAM HDTV and Grand Alliance HDTV receivers. Since September 1996, he has been with the School of Electronics Engineering, KNU as a Professor. He worked as a courtesy Associate Professor in the department of electrical and computer engineering, University of Florida in 2004. He was director at the center of Digital TV and Broadcasting in the Institute for Information Technology Advancement (IITA) from July 2006 to July 2008. His main research interests include artificial intelligence in self-driving cars and communication systems.