Illumination-Robust Foreground Extraction for Text Area Detection in Outdoor Environment

Jun Lee[1], Jeong-Sik Park[2], Chung-Pyo Hong[3] and Yong-Ho Seo[1]
[1] Department of Intelligent Robot Engineering, Mokwon University
Daejeon, Republic of Korea
[e-mail: leetrue123@nate.com, yhseo@mokwon.ac.kr]
[2] Department of Information and Communication Engineering, Yeungnam University
Gyeongsan, Republic of Korea
[e-mail: parkjs@yu.ac.kr]
[3] MC Division, LG Electronics
Seoul, Republic of Korea
[e-mail: hulkboy@yonsei.ac.kr]
*Corresponding author: Yong-Ho Seo

Abstract

Optical Character Recognition (OCR) that has been a main research topic of computer vision and artificial intelligence now extend its applications to detection of text area from video or image contents taken by camera devices and retrieval of text information from the area. This paper aims to implement a binarization algorithm that removes user intervention and provides robust performance to outdoor lights by using TopHat algorithm and channel transformation technique. In this study, we particularly concentrate on text information of outdoor signboards and validate our proposed technique using those data.

1. Introduction

Optical Character Recognition (OCR) is a technology to recognize text area in videos or images taken by digital devices includinng camera and convert them into letters readable by machines. OCR is one of the main research areas of artificial intelligence or computer vision. Advanced countries were the first to begin OCR studies, but development of OCR technology has been mainly concentrated on English characters.

According to languages, letters or characters represent different features in terms of form or shape. For this reason, the conventional OCR techniques adopted for English characters may not be applicable to other languages. This paper focuses on an advanced technique for OCR to be adopted for Korean language.

In Republic of Korea, a software for recognizing characters of Hangul (name of Korean language) was firstly introduced in early 1990s, giving unsatisfactory performance. Since then, OCR technologies for Hangul have been steadily studied, but most of them had limited applications because they were operated on workstation-level machines with excellent processing capacity. However, as personal computer capacity improves and even handheld devices such as smartphones provide reliable processing capability, researchers have extended research areas of OCR, drawing a variety of applications using OCR [1].

In Section 2, we summarize the conventional studies on OCR. Section 3.1 and Section 3.2 explain an algorithm for illumination-robust text area detection in region of interest (ROI) and classification rules for combining characters of Hangul, respectively. In Section 3.3, a binarization algorithm for text area detection in outdoor environment is proposed. Section 4 demonstrates experiments on IDCAR datasets and results. Finally, Section 5 concludes this paper.

2. Related Work

Elementary technologies on OCR have been actively studied for the past years. Since character recognition, however, is higly dependent on language information, it was difficult to apply the advanced technology based on English to other lanuages including Hangul in which a syllable is formed by combination of characters. For this reason, separate studies for Hangul character recognition have been continuously demanded in Korea. Especially, with penetration of smartphones as a personal digital device, diverse OCR applications have been developed by utilizing smartphone-embedded camera functions. For its wide variety of usefulness such as auto-interpretation and augmented reality, more excellent technologies are continuously demanded.

In the character recognition process, text extraction is a fundamental phase for obatining reliable performance of character recognition. To extact text areas, essential characteristics of character elements existing in natural scenes are utilized [2]; the inner color of character elements is homogeneous and character color is well distinguishable from surrounding background colors. However, text area detection in outdoor environments has several difficulties. One of them is diversely different shooting environments of natural scenes. Most visual contents taken by camera do not provide homogeneous environments including brightness. Such dynamic shooting environments bring various problems such as light and shadow, complicated background and reflection according to the quality of medium upon

which characters are placed, undermining the precision of character detection [3][4]. In this situation, this study seeks to make up for the potential performance-undermining elements in the outdoor environment. To this end, this study acquires user intervention to present recognition area and suggests an illumination compensation technique and binarization pattern selection technique [5]. And for the recognition of Hangul, combination characters, this study proposes a technique to divide text areas based on syllables. For more in-depth study, an algorithm is proposed herein, which proactively performs global binarization based on the candidate text area distribution and extracts text areas without user intervention.

3. Foreground Extraction for Text Area Detection in Outdoor Environment

3.1 Illumination-Robust Text Area Detection in ROI

Text area detection is a fundamental task for reliable performance of character recognition. However, text area detection in outdoor environments has several difficulties drawn by the diverse different shooting environments of natural scenes. Most videos taken by camera do not provide homogeneous environment including brightness. Such dynamic shooting environments bring about various problems such as light and shadow, complicated background and reflection according to the quality of medium upon which characters are placed, undermining the precision of character detection. Figure 1 shows an example.
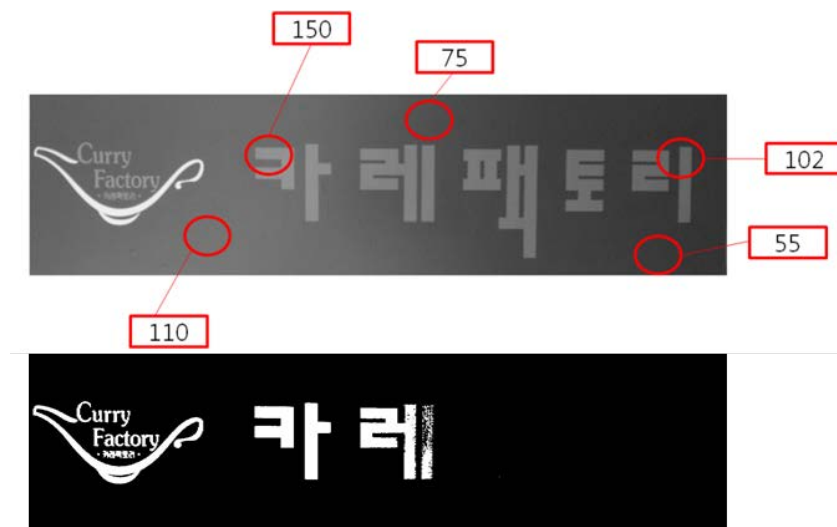
Fig. 1. Image (up) representing the average pixel value of a random area in non-constant illumination video and image (down) of binarization based on the average value

To compensate such distribution imbalance of illumination, this study utilized L channel of LAB Color Model along with Median Filter to reduce illumination interference. LAB Color Model uses image brightness and color information to indicate colors. Thus, lightness information is easily separated. Each channel represents L (lightness), A (Red/Green color value), and B (Yellow/Blue color value). In general RGB videos, the conversion equation is as follows [6].

$$L^* = 116\mathrm{f}\left(\frac{y}{y_n}\right) - 116$$
$$A^* = 500\left[f\left(\frac{x}{x_n}\right) - f\left(\frac{y}{y_n}\right)\right]$$
$$B^* = 200\left[f\left(\frac{y}{y_z}\right) - f\left(\frac{z}{z_n}\right)\right] \tag{1}$$

f(s) represents $s^{1/3}$ where $s > 0.00856$ and represents $7.787s + 16/116$, elsewhere.

Here, the color difference between A and B in LAB Color Space, ΔE, can be estimated as follows.

$$\Delta\mathrm{E} = \mathrm{m}\sqrt{(L_2^* - L_1^*)^2 + (A_2^* - A_1^*)^2 + (B_2^* - B_1^*)^2} \tag{2}$$

The median value-based filtering technique, Median Filter, is applied to the L channel produced by the equation above to acquire a filtered image. In this study, the acquired images are regarded as light distribution status. The final illumination compensation can be realized through the association of difference between the original video and lighting channel. And it can be acquired through the calculation of difference with the gray-conversion of original image. The following **Fig. 2** shows the image filtered by Median Filter in L channel is regarded as light distribution status and **Fig. 3** shows the illumination compensated image acquired based on the image difference between gray-converted original image and illumination image. **Fig. 4** shows the overall flow of proposed illumination compensation and binarization algorithm.

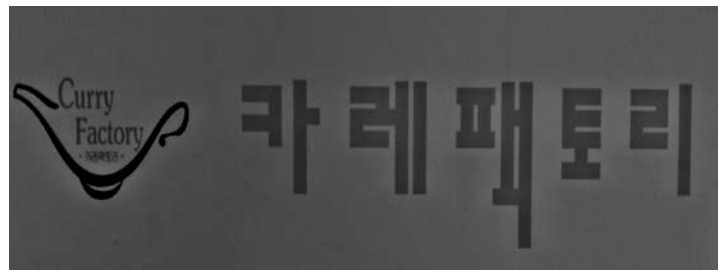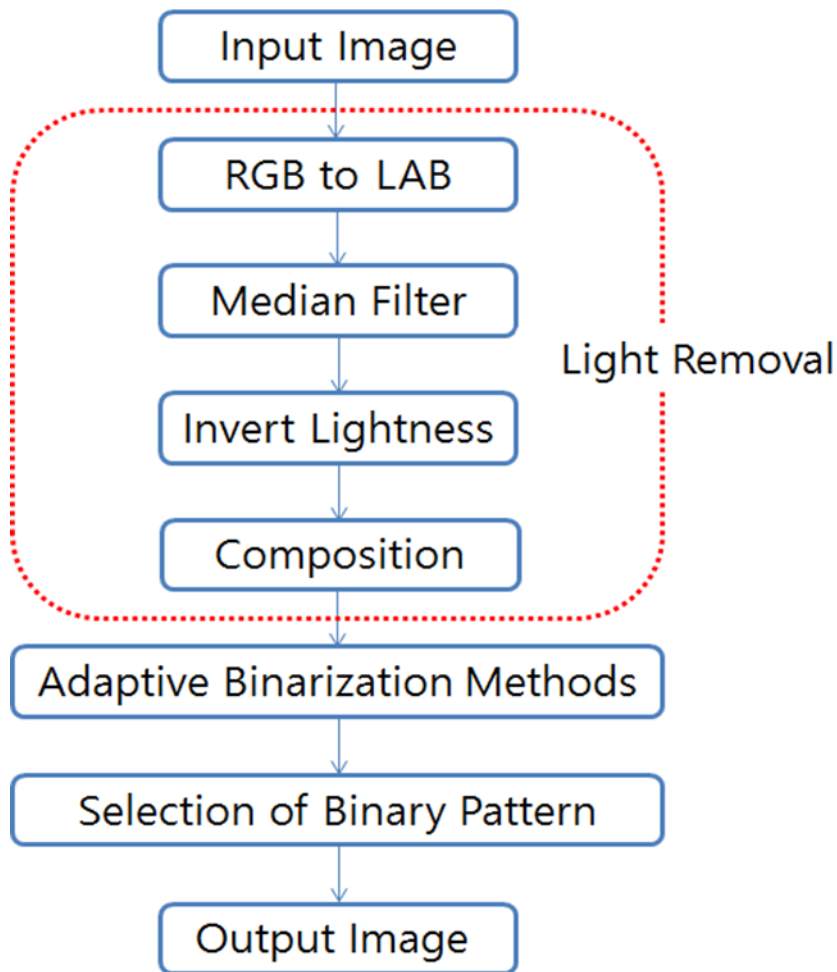**Fig. 2.** The image filtered by Median Filter in L channel is regarded as light distribution status.

Fig. 3. Illumination compensated image acquired based on the image difference between gray-converted original image and illumination image

Fig. 4. Flow of proposed illumination compensation and binarization algorithm

To effectively compensate uneven light distribution in natural scenes and noise other than text area, this study adopts a threshold value changed based on the brightness mean and variance of each part under the Niblack binarization algorithm. Niblack binarization equation is as follows [7].

$$T = m + k \times s \qquad (3)$$

Here, k is a constant to be set by a user. Normally 0.02 is recommended. **m indicates the size of area.** The threshold value, T, is not sensitive to the size of partial area but should be determined properly according to each application case. To determine the value for the given equation, mean and variance should be repeatedly calculated in the partial area and the calculation time is dependent on the area size.

$$T = m + k \sqrt{\frac{1}{N} \sum (P_i - \mu)^2}$$

$$= m + k\sqrt{\frac{\sum P_i^2}{N} - \mu^2}$$

$$= m + k\sqrt{\sigma^2}$$

(4)

In this equation, N, μ, and σ^2 represent the number of entire pixels, the mean and variance of p_i, respectively. Finally, Niblack binarization algorithm is applied to the image processed by the proposed lighting removal technique to acquire a binary image with separated foreground and **background**. **Fig. 5** shows the Niblack binarization algorithm-applied image after lighting removal.

Fig. 5. Niblack binarization algorithm-applied image after lighting removal

3.2 Syllable-level classification in Hangul text area

Alphabet has a total 26 characters from A to Z and, if both capital and small letters are considered, there are 52 alphabet characters. On the other hand, Hangul has 19 consonants and 21 vowels to make the total of 40 letters but, if combined, they can make up to approximately 540,000 different letters. Among them, the number of letters available in digital expression is 2,350, having over 45 times more classes than alphabet. For this reason, it has limitations on applying English character-specialized recognition technology to Hangul. It is necessary to classify each text area element into syllables before character recognition. In this study, for the syllable classification, the characteristics of Hangul are re-interpreted according to its types as combination characters. As a result, among six types of Hangul, four were found to have each element in vertical component and remaining two types were found to have each consonant/vowel element in horizontal relationships. Based on such a characteristic, the six types of Hangul in the linguistic dimension are re-classified into two types. **Fig. 6** shows the six types of Hangul in the linguistic dimension were re-classified into two types according to the vertical/horizontal relationships
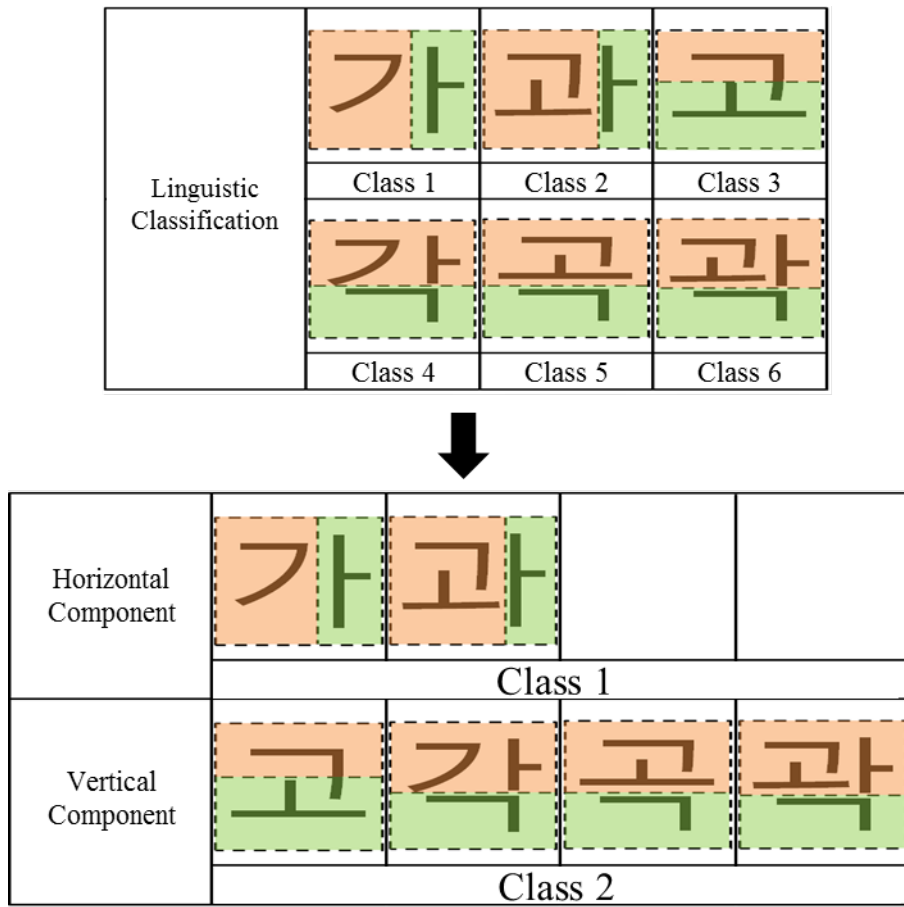
Fig. 6. The six types of Hangul in the linguistic dimension were re-classified into two types according to the vertical/horizontal relationships.

To classify the types according to components' horizontal/vertical relationships, morphology method is applied in this study. First, to separate the four types where syllable consonant and vowel are in vertical arrangement, a vertical blur is applied on the binary image, as shown in **Fig. 7**.

$$A \oplus B = A^c \ominus (-B^c) \tag{5}$$

Fig. 7. Vertical blur-applied binary image

Then, to divide syllable candidate areas, a labeling algorithm is employed as an area division method for binary images. The sets of non-neighboring pixels are classified and they are

designated as each syllable candidate. After labling, a set of pixels gathered together under one single lable is regarded as the candidate of one syllable or syllable-consisting component, as shown in **Fig. 8**.

Fig. 8. Labeling algorithm application to Fig. 4 image to group neighboring pixels

Following processes determine which class each acquired label falls under;

1. Calculate the average, T, of areas of every label ($L_0 \sim L_n$).
2. If label area is smaller than T × 0.5, classify the corresponding label into Class 1.
3. Repeat T value updating based on the average of remaining label areas.
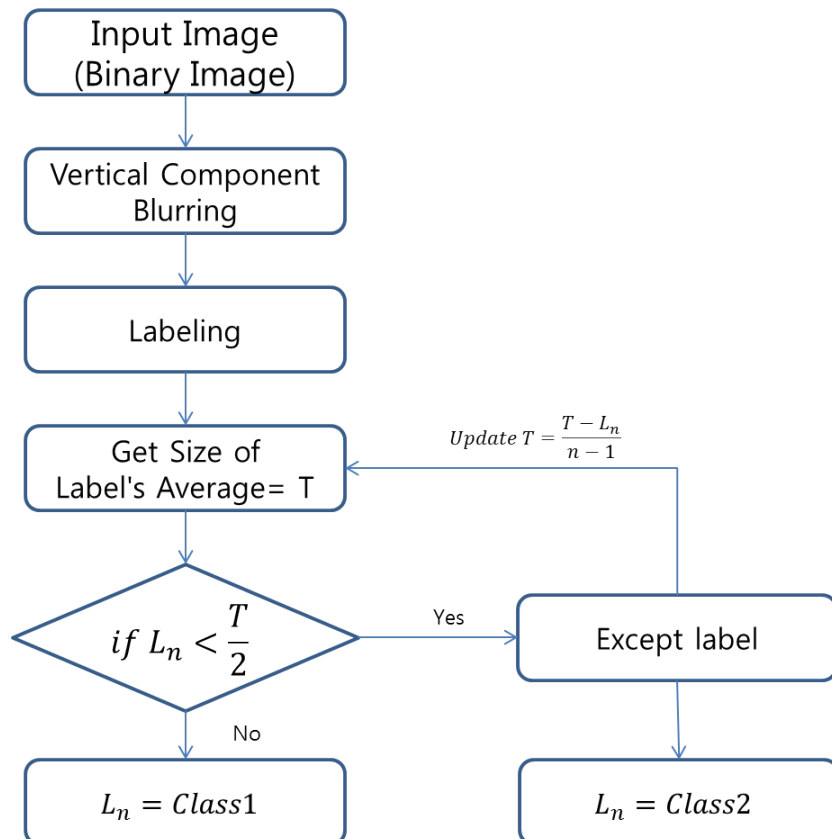4. If T value does not increase, classify the rest of the labels into Class 2.
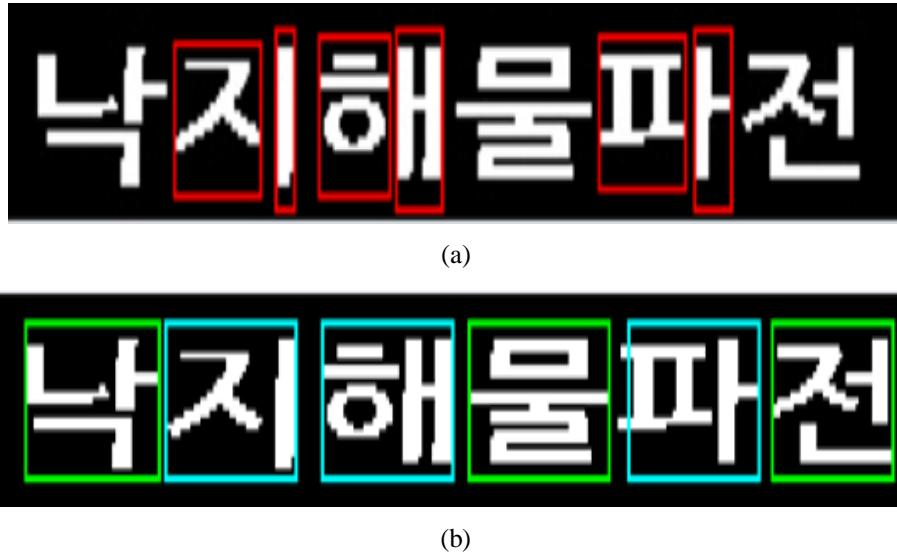

Fig. 9. Flow of the proposed syllable classification algorithm

Finally, based on the 2 class labels classified above, the classified syllables are obtained. The lable candidates under Class 2 are integrated into a single lable by utilizing T value, the size of Class 1.

(a)

(b)

Fig. 10. Labels classified into Class 2 (a) and final results of syllable classification (b)

3.3 Binarization Algorithm for Text Area Detection in Outdoor Environments

Our proposed text detection technique shows a high recognition rate, but it requires users to set ROI (Region of Interest). In other words, the proposed method becomes unoperable in proactive platforms such as robot where user intervention is difficult [8]. In order to address this problem, this study proposes another algorithm to conduct global binarization of text area outdoors without user intervention.

To perform binarization in outdoor environments, illumination-robust images need to be acquired. In this paper, the process in 3.1 is employed to suggest a single-channel illumination-robust image. The proposed image, K, is obtained as follows.

$$f(x, y) = (L, A, B)$$

$$total = (L + A + B)$$

$$A' = \frac{A}{total} \times 255$$

$$B' = \frac{B}{total} \times 255$$

$$K(x, y) = \frac{(A' + B')}{2}$$

(6)

Active binarization algorithm is upon the Top Hat algorithm [9], which is a morphology technique to extract a specific interested object from an input image. In the gray-converted original image, opening or closing operation is performed. Then, the difference is calculated with the gray-converted original image to extract the interested object. The method is frequently used to extract number plates of vehicles. However, the opening or closing calculation needs to be selected depending on the relative brightness of the interested objects in relation to the background. Thus, it is not appropriate for sign images in outdoor environments that convey diverse levels of different colors and brightness between interested object and background, unlike number plate cases [9].
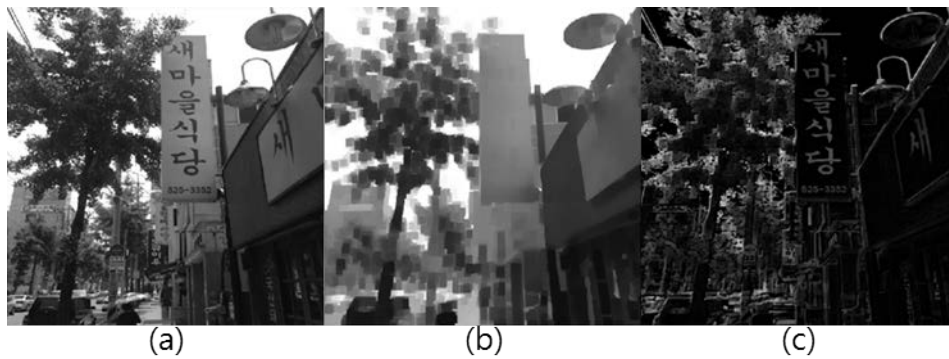
Fig. 11. Gray image (a), Morphology image (opening) (b), and Subtracted image (c)

If a Top Hat algorithm-applied image is processed under binarization, and then labeling is conducted, the performance may plummet in outdoor environments due to numerous text candidates as shown in **Fig. 12**.
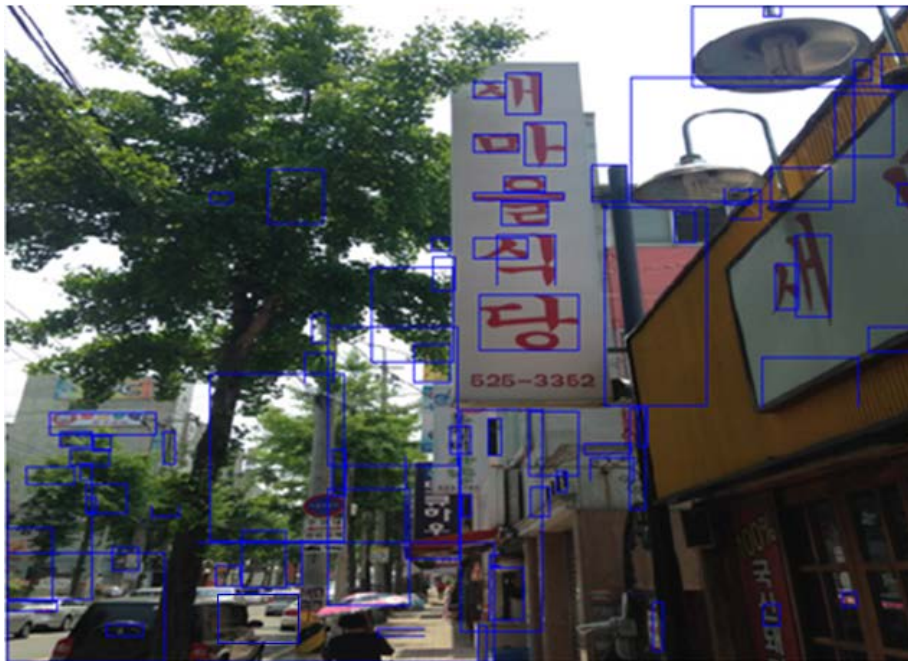
Fig. 12. Top Hat algorithm-based text area candidate detection

For this reason, this study proposes a histogram filtering technique based on histogram characteristics, in order to select text candidates more precisely among a variety types of text candidates observed in natural scenes. The histogram technique used herein is based on Otsu's binarization algorithm. Assuming that the ROI histogram is a bimodal type, we find valley points and classify them into two classes. The valley σ^2 can be represented as the sum of within-class variance, $\sigma_\omega^2(t)$, and between-class variance, $\sigma_b^2(t)$. When each within-class standard deviation satisfies 68-95-99.7 rules, the corresponding area is regarded as a significant candidate. Binarization is performed finally based on the between-class variance between two calsses [10].

$$\sigma_\omega^2(t) = \omega_1(t)\sigma_1^2(t) + \omega_2(t)\sigma_2^2(t)$$
$$\sigma_b^2(t) = \sigma^2 - \sigma_\omega^2(t) = \omega_1(t)\omega_2(\text{t})[\mu_1(t) - \mu_2(t)]^2 \tag{7}$$

Chosen from the candidates
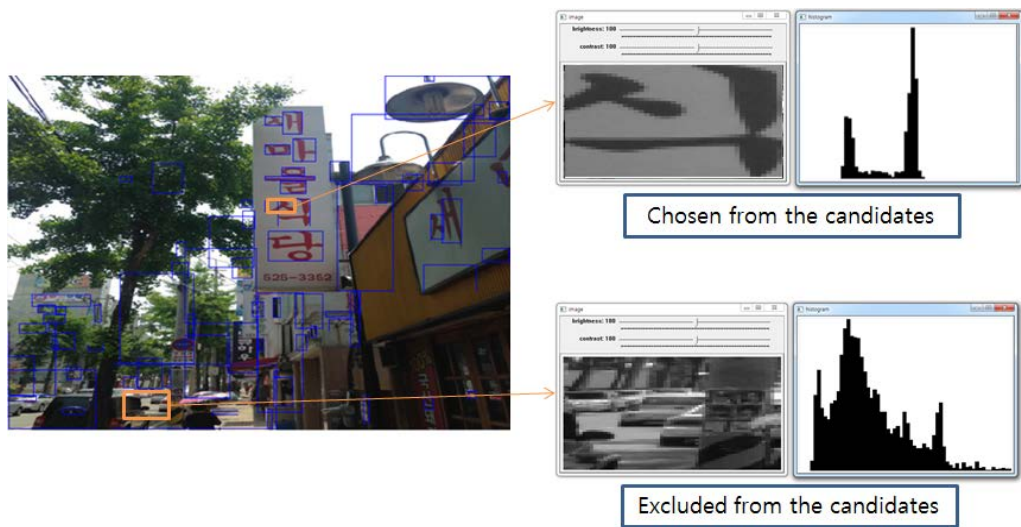
Excluded from the candidates

Fig. 15. Top Hat algorithm-based text area candidate detection

(a)   (b)

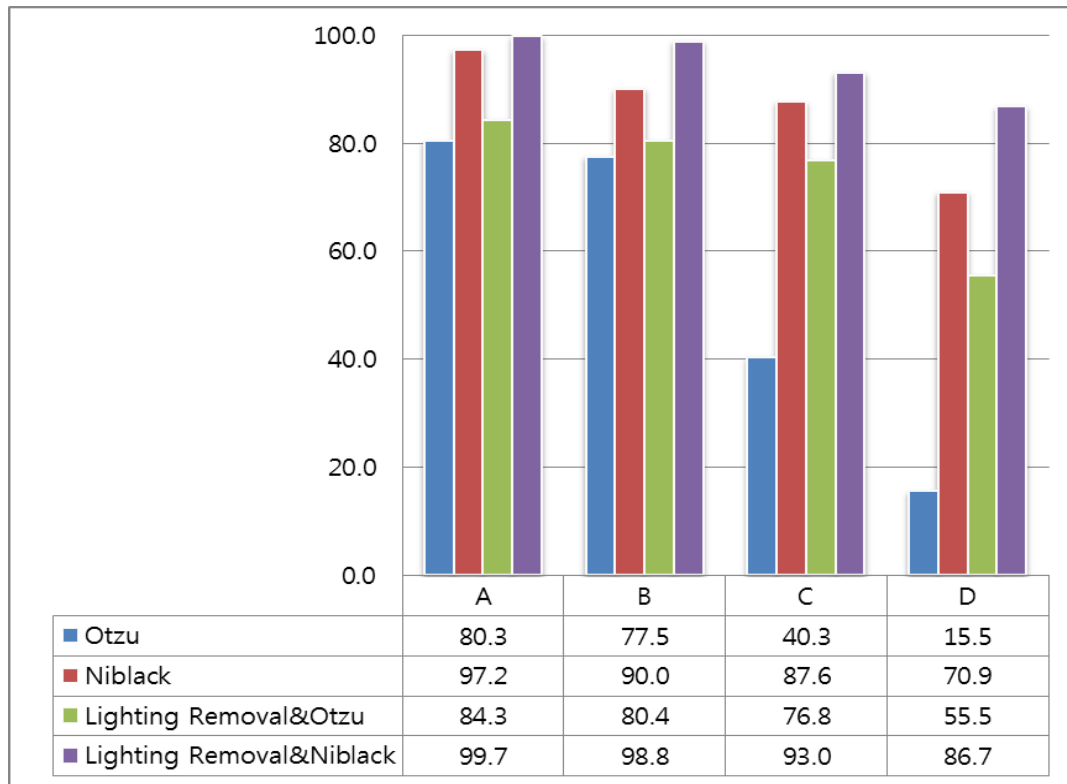**Fig. 16.** Pre-reduction candidate (a) and post-reduction candidate (b)

Fig. 17. Test result of binarization algorithm for text area detection in outdoor environments

4. Experimental Results and Analysis

To verify the proposed approach, we performed experiments upon image database that is the high-resolution image dataset taken by smartphones. This dataset was originally made by the Electronics and Telecommunications Research Institute of Republic of Korea. Most the image subjects contain shop signs and menus taken in outdoor environments. A total of 113 randomely-selected database images were utilized for experiments. We performed several experiments, separately, including illumination compensation and binarization in ROI, syllable classification, text area binarization in global image, and text area extraction from global image.

First of all, the ROI binarization algorithms were verified with global binarization, Niblack adaptive binarization, the proposed illumination compensation algorithm-based global binarization, and illumination compensation algorithm-based adaptive binarization. Experiments were conducted on the database consisted of four classes depending upon the levels of illumination interference. As a result, the adaptive binarization algorithm applying the proposed illumination compensation algorithm showed the highest performance of 86.7~100%. In particular, in an experiment on dataset of the most severe illumination interference, the Niblack adaptive binarization algorithm recorded the second highest performance with 70.9%, but our proposed method showed 15.8% of relative improvement compared to the algorithm.

	A	B	C	D
■ Otzu	80.3	77.5	40.3	15.5
■ Niblack	97.2	90.0	87.6	70.9
■ Lighting Removal&Otzu	84.3	80.4	76.8	55.5
■ Lighting Removal&Niblack	99.7	98.8	93.0	86.7

Fig. 17. Experimental results of illumination-robust text area detection algorithm in ROI

To classify Hangul syllables, the existing characteristics of each Hangul type were reinnterpreted. And the conventional six types categorized according to last consonant status were re-organized into new types according to the vertical/horizontal relation of consonant and vowel. Then, the vertical morphology technique and filtering technique were employed. Printing letters photographed at the front were evaluated. A hundred text area images were randomely selected from 60 images. In this experiment, our approach recorded 94%.

Table 1. Results of syllable classification evaluation

Number of Image	Number of Test	Test Result(%)
60	100	94

5. Conclusion

In this study, we porposed an active binarization algorithm to avoid user intervention in outdoor environments. The proposed method is based upon the text area extraction algorithm based on channel transformation-based binarization and Top hat algorithm, a morphology technique to extract a specific object.

In the experiments, the performance of binarization algorithm achieved 87.6% and that of the text area extraction algorithm recorded 74.3% when the noise was invoked lower than 10% of text. Concerning the performance evaluation, the binarization algorithm performance was determined mostly in the process of Top Hat algorithm-based text area selection and the process of determining binarization reversal.

Since the text area extraction algorithm only detects the color of characters with the strongest probability in the character color candidate extraction process, it showed much frequent errors when there were two or more colors in subject characters. In the future work, we will imporve our proposed algorithm to realize active text area detection and character recognition in a platform environments such as robot.

Acknowledgement

References

[1] D. Doermann, J. Liang, and H. Li, "Progress in camerabased document image analysis," in *Proc. of Document Analysis and Recognition, in Proceedings, Seventh International Conference*, pp. 606–616, 2003. Article (CrossRefLink).

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

[3] Cong Yao, Xiang Bai, Wenyu Liu, Yi Ma, Zhuowen Tu, "Detecting texts of arbitrary orientations in natural images," in *Proc. of Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference*, pp.1083-1090, 2012. Article (CrossRefLink).

[4] N. Ezaki, M. Bulacu, and L. Schomaker, "Coarse-to-Fine Description for Fine-Grained Visual Categorization," *IEEE Transactions on Image Processing*, vol. 25, pp. 4858–4872, 2016. Article (CrossRefLink).

[5] Sergey Milyaev, Olga Barinova, Tatiana Novikova, Pushmeet Kohliy, Victor Lempitsky, "Image binarization for end-to-end text understanding in natural images," in *Proc. of International Conference on Document Analysis and Recognition*, pp. 128-132, Aug 25-28, 2013 Article (CrossRefLink).

[6] N. Otsu, "A threshold selection method from gray level histograms," *IEEE Trans. Systems, Man and Cybernetics*, vol. 9, pp. 62–66, 1979. Article (CrossRefLink).

[7] Khurram Khurshid, Imran Siddiqi, Claudie Faure, "Comparison of Niblack inspired binarization methods for ancient documents," in *Proc. of Document Recognition and Retrieval XVI*, January 18, 2009 Article (CrossRefLink).

[8] L. Neumann and J. Matas, "Real-time scene text localization and recognition," *CVPR*, pp. 3538–3545. 2012. Article (CrossRefLink).

[9] Gatos, B., et al, "Text detection in indoor/outdoor scene images," in *Proc. of First Workshop of Camera-based Document Analysis and Recognition*, 2005.

[10] N. Ezaki, M. Bulacu, and L. Schomaker, "Text Detection from Natural Scene Images: Towards a System for Visually Impaired Persons," in *Proc. of In International Conference on Pattern Recognition*, pp. 683-686, 2004. Article (CrossRefLink).

Jun Lee received B.S. and M.S. degrees from the Department of Intelligent Robot Engineering, Mokwon University, Republic of Korea, in 2013 and 2015, respectively. He is currently a Ph.D candidate of Intelligent Robotics Lab, Mokwon University. His research interests include the areas of computer vision, activity recognition and vision-based surveillance system.

Jeong-Sik Park received his B.E. degree in Computer Science from Ajou University, Republic of Korea, in 2001 and his M.E. and Ph.D. degree in Computer Science from KAIST (Korea Advanced Institute of Science and Technology) in 2003 and 2010, respectively. From 2010 to 2011, he was a Post-Doc. researcher in the Computer Science Department, KAIST. And he was an assistant professor in the Department of Intelligent Robot Engineering, Mokwon University, from 2012 to 2013. He is now an assistant professor in the Department of Information and Communication Engineering, Yeungnam University. His research interests include signal processing, voice interface for human–computer interaction and machine intelligence

Chung-Pyo Hong received the B.S. and M.S. degree in computer science from Yonsei University, Seoul, Korea, in 2004 and 2006, respectively. In 2012, he received the Ph.D. degree in computer science from Yonsei University, Seoul, Korea. He is currently a Senior Researcher at the LG Electronics. His research interests include mobile computing, and ubiquitous service platform.

Yong-Ho Seo received his B.S. and M.S. degrees from the Department of Electrical Engineering and Computer Science, KAIST, in 1999 and 2001, respectively. He also received a PhD degree at the Artificial Intelligence and Media Laboratory, KAIST, in 2007. He was an Intern Researcher at the Robotics Group, Microsoft Research, Redmond, WA in 2007. He was a consultant at Qualcomm CDMA Technologies, San Diego, CA in 2008. He is currently a Professor of the Department of Intelligent Robot Engineering, Mokwon University. His research interests include humanoid robot, human-robot interaction, robot vision and wearable computing.