

Application on character recognition system on road sign for visually impaired: case study approach and future

Jaejoon Kim

School of Computer and Communication, Daegu University, South Korea

Article Info

Article history:

Received Apr 7, 2019

Revised Sep 18, 2019

Accepted Sep 27, 2019

Keywords:

Character recognition system

MSER

OCR

Visually impaired

Wearable device

ABSTRACT

Many visually impaired people worldwide are unable to travel safely and autonomously because they are physically unable to perceive effective visual information during their daily lives. In this research, we study how to extract the character information of the road sign and transmit it to the visually impaired effectively, so they can understand easier. Experimental method is to apply the Maximally Stable External Region and Stroke Width Transform method in Phase I so that the visually impaired person can recognize the letters on the road signs. It is to convey text information to the disabled. The result of Phase I using samples of simple road signs was to extract the sign information after dividing the exact character area, but the accuracy was not good for the Hangeul (Korean characters) information. The initial experimental results in the Phase II succeeded in transmitting the text information on Phase I to the visually impaired. In the future, it will be required to develop a wearable character recognition system that can be attached to the visually impaired. In order to perform this task, we need to develop and verify a miniaturized and wearable character recognition system. In this paper, we examined the method of recognizing road sign characters on the road and presented a possibility that may be applicable to our final development.

Copyright © 2020 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Jaejoon Kim

School of Computer and Communication,

Daegu University,

201 Daegudae-ro, Gyeongsan, Gyeongbuk, 38453, South Korea.

Email: jjkimisu@daegu.ac.kr

1. INTRODUCTION

There are many people with visual disabilities in the world as well as people reside in Korea. Auxiliary devices have been actively developed to provide visually impaired persons with accurate information on road signs during walking, which can be proven as crucial parts of everyday social activities. Even more, various devices such as products that read or provide directions for blind have been actively developed in the past.

Various methods are currently in development to automatically extract the region surrounding a character when that character is detected to be in the presence of a visually impaired person [1-5]. The method consists of mainly first recognizing the character string in its region and then sending the scanned information back to the user. Due to artificial intelligence technology, additional features can include a dictionary function in case the user does not understand the meaning of the sign in question. In recent years, various techniques using artificial intelligence in image recognition, natural language processing, and natural language generation have been developed to help people with visual impairments to maintain a comprehensive and productive life [6].

In this paper, we tried to develop a variety of wearable assistive aids for the visually impaired to eliminate the inconvenience of walking or real life. In order to do this, in Phase I, Maximally Stable External Region (MSER) and Stroke Width Transform (SWT) features were applied to character recognition of road signs and pedestrian signs [7, 8]. In Phase II, we tried to design a text recognition system based on TTS (Text-to-Speech) converter [9] that transmits the text recognition information to the visually impaired.

This paper proposes a plan for the visually impaired to meet these developments and to present the applicable possibilities. Section 2 briefly discusses the related works for character string recognition technology currently in development for the visually impaired and characteristics of information guiding system. Section 3 describes experimental implementation of character recognition for our proposed system. Sections 4 and 5 conclude the paper by discussing experimental results, future works and other possible solutions.

2. RELATED WORK

2.1. Characteristics and trend of information guiding system

Blind people are divided into low vision and the blind. Low vision refers to vision problems that cannot be improved by medical or optical means due to birth defects or acquired eye diseases. According to a report of Yonsei University Medical Center, the suicide rate of the low vision is twice that of the normal person, and the psychological stress due to the low vision is large, and the need for an auxiliary device to help the low vision is increasing [10]. In addition, a report from the UN estimates that the world’s population age 60 or older will increase from 11% in 2012 to 22% by 2050 as shown in Figure 1. Especially in North America and Europe, the distribution of population aged 60 years or older is reported to be significantly higher than the average [11]. In the point of world market, Figure 2 indicated that the global elderly and disabled assistive devices market was valued at US\$ 14,109.1 million in 2015 and is expected to expand at a CAGR of 7.4% during the forecast period (2016 – 2024) [12].

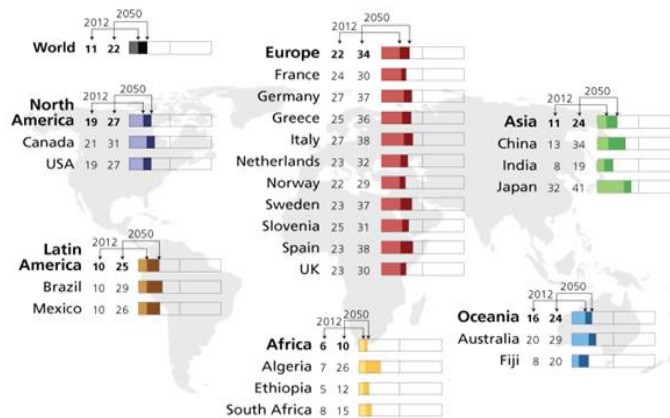


Figure 1. The statistics of UN, gendered innovations (2012 vs. 2050) [11]

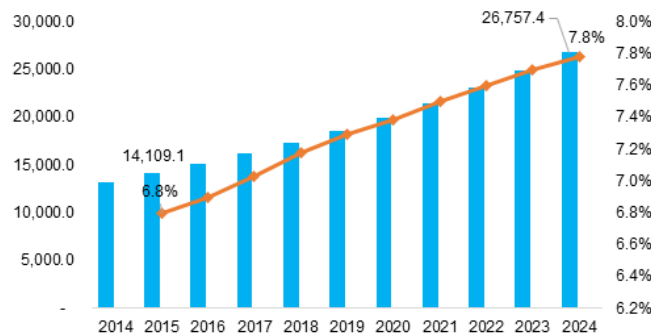


Figure 2. Global elderly and disabled assistive devices market size and forecast (unit: US\$ Million) [12]

2.2. Character information recognition system

To allow a visually impaired person to recognize a character, our application must be able to find the character region in which the sign or the message is written. The most well-known method used for binarization is a global binarization which is based on a single threshold [13-15]. However, these methods have characteristics such as performance issues with the overall brightness of the input document image are not constant.

Another method used for binarization is the Maximally Stable External Region (MSER) algorithm, which is used in a more robust manner of finding a region with a small rate of change while changing the threshold value in order to distinguish characters in both bright and dark regions. The character area is generally easy to distinguish from the background and has a constant brightness value. This characteristic of the character region is suitable to be detected by the MSER method [7, 8]. Various applications have been developed by various companies, all of which attempt to extract a character string by extracting horizontal and vertical boundary components from the input image. These methods basically either find the portion with the greatest boundary component or extract a character string that is irrespective of misjudgment or distortion.

The ABBYY FineReader Engine, OCR SDK (character recognition development toolkit) supplied by Retia, was applied to the development of a 'print-to-speech converter', which is a device for recognizing printed characters for visually impaired and converting them into voice [16-18]. ABBYY [19] developed ABBYY FineReader Engine with a goal of a stable 'print-to-speech' converter for the visually impaired. The OCR key technologies of ABBYY FineReader Engine have several main features such as:

- Korean, Chinese, English, numbers and special characters support
- Vertical reading, Multi reading
- Batch scanning, batch conversion possible
- It can save word, text file etc.

By utilizing this engine, the visually impaired user can convert the desired document to speech through this device. While in the past books, receipts and other readable documents could not be read without Braille documents, the engine now makes it possible to directly scan a desired document or image by using the 'sound' with OCR technology and convert it so the user can hear the by voice. The 'finger reader' [20] product developed by MIT works by placing a finger near the book. A high-resolution camera is used to scan the characters corresponding to the finger's touch while also reading them out loud after being scanned. After several years of research, OrCam, an Israeli company, is now developing new products that use facial recognition technology to scan not only characters but also various colors [21]. There are also 'smart glasses' for people with partial vision loss. These 'smart glasses' work by viewing images in 3D glasses. However even though these glasses have been proven useful reducing inconvenience of the visually disabled, they are still very expensive. Therefore there will be difficulties in making them commercially available for everyone. Figure 3 shows the number of people in Korea with sight disabilities is around about 250,000 people as of 2016. Unfortunately, the number of people with visual disabilities is growing larger each year by more than 10%. In addition, the types of disabilities are described in Figure 4.

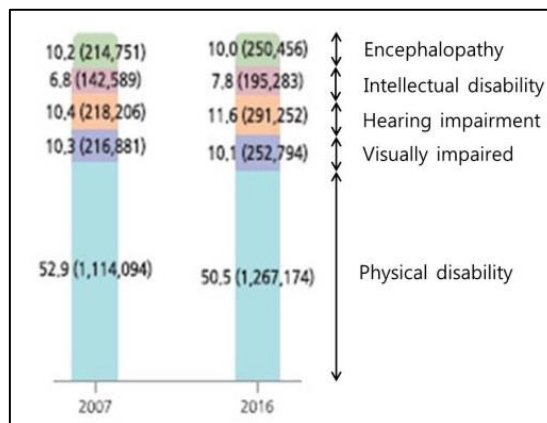


Figure 3. The statistics of disabilities in Korea, 2007 and 2016 [22]

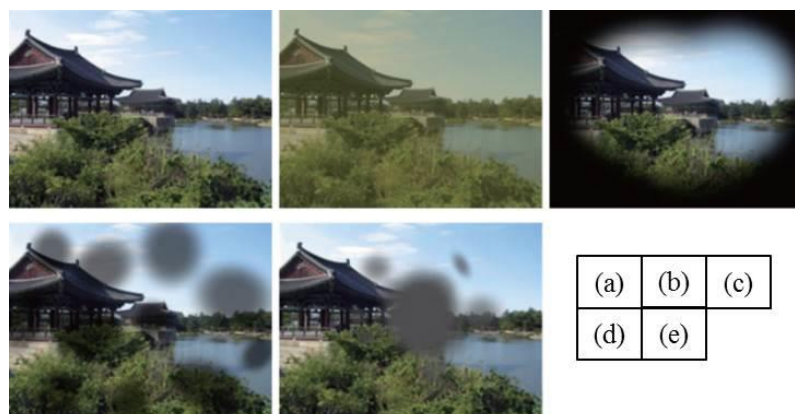


Figure 4. The types of visually impaired. (a) Scenery seen with normal eye, (b) Scenery seen with eyes with cataracts, (c) Scenery seen with eyes with glaucoma, (d) Scenery seen with eyes with diabetic retinopathy, (e) Scenery seen with eyes with macular degeneration [23]

3. EXPERIMENTAL IMPLEMENTATION

One of the most important parts of the application is to provide information that is only necessary in order to never confuse the user. Therefore, the application must be able to select only the relevant area amongst the other ‘background noise’. For example, the application must be able to distinguish between a road sign and the other natural background surrounding the sign. In this paper, we aim to detect the text area in the image provided with the visually impaired and to recognize the information provided in the area as a character and deliver it to the visually impaired. In order to implement our goal, we used MSER and OCR methods for detecting information area and text detection as a starting point of research. Based on the Phase I of the research, we aim to develop a wearable device that provides road information about the visually impaired by utilizing the TTS (Text-to-Speech) function of a typical smartphone.

3.1. Feature extraction with MSER

Information on the road signs do not contain as much textual information when compared to books or newspapers. As a method of the visually impaired to arrive at their desired destination, they had to previously rely the usage of a guidance stick. While it may be practical and has been used as from the beginning of blind people, a different method can be developed. These days it is possible to detect the text in an image using no more than a smartphone with a camera. By using pixel spans as nodes, blobs can be generated which can detect text in an image more quickly and accurately. In addition, as the usage of smartphone increased very rapidly, an optical character recognition application that recognizes a character of an image captured by a camera mounted on a smart phone and displays it on a display can be distributed in the near future and prove more practical compared to a guidance stick.

The most important step is to detect the blob which does a candidate region exists in the image that is most likely to contain the text of the road sign. The SWT algorithm [24, 25] is a method of detecting text by determining which region has little change in the shape and thickness. When the parameters are given, the algorithm detects the text within the image using those parameters. The main advantage of SWT are that it can detect text in an image without a separate learning process, but it also has a drawback in which it takes a lot of time to detect text in an image because of its complicated operation. The MSER algorithm for robust text detection is widely used as an algorithm for detecting blob as an aggregate region of pixels different from surrounding pixels in intensity. The MSER algorithm has the advantage that it can detect the blob faster than the SWT algorithm, but it has a disadvantage that the accuracy is somewhat lacking, such as detecting the tickle or noise by blob.

3.2. Application procedure

Figure 5 is an overall block diagram that shows the process of our application after capturing the image and extracting the text. In this paper, we aimed to test the design possibility of an assistive device which can be helpful for the visually impaired through Phase I process in two stages. The first step captures images of a smartphone or future terminal device that the blind has and then removes non-text regions using MSER. The MSER extraction implements the following steps:

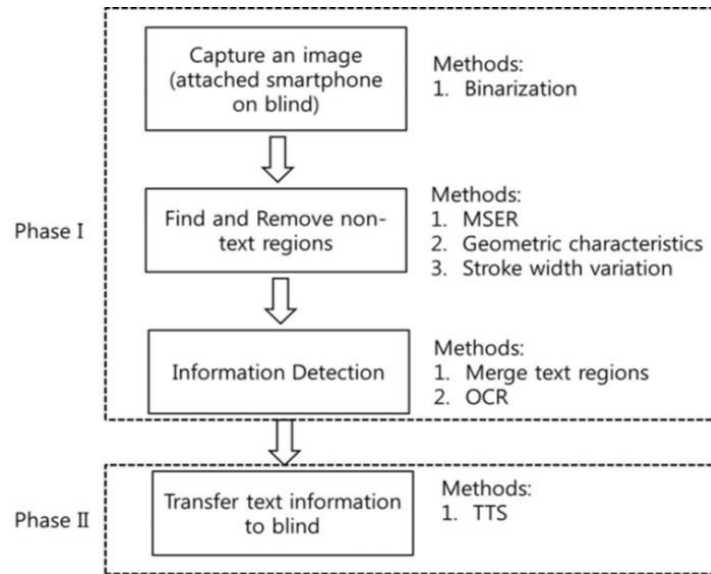


Figure 5. The overall block diagram

- a. Sweep threshold of intensity
- b. Extract connected components
- c. Find a threshold on maximally stable
- d. Approximate a region with an ellipse
- e. Keep region descriptors

One of the most important features of the MSER algorithm is that it detects most text in an image but also extracts a stable region, not a text region. To make the algorithm suit our needs, we removed non-text areas using parameters that distinguish words by their geometrical properties within the image. In addition, SWT can be used to distinguish more sophisticated non-text areas. In this paper, we applied this method to find and remove non-text areas. In order to merge individual text areas to extract information on a single word or a text line after finally confirming the text area, a neighboring text area is searched to form a bounding box of these areas, and finally, the information is extracted. In Phase II, the final detected text information is converted into voice information on the visually impaired to be transmitted or applied to future applicable and wearable devices.

4. RESULT AND DISCUSSION

4.1. Phase I

In our experiment, we used a typical image of the road sign which the visually impaired can easily encounter on the road. The left column of Figure 6 shows test data set. These images were sorted in order to make test conditions more difficult. When comparing the amount of text to the whole image there is not much text contained in the image in the first place. Therefore, these images were ideal in testing if our proposed method of combining MSER and SWT to extract the necessary information. These test conditions were also important because the end goal of our application is to distribute it so that it can be used during everyday activities. Thus, we made sure that the images were selected very carefully in order to simulate what the user might encounter while using the application.

In Figure 6, we show the results of applying MSER, geometric characteristics, and then applying SWT to find non-text areas in captured images using mobile phones of the visually impaired. These results are shown on right column of Figure 6 and Table 1 shows the final detected character information of each test image. In our test case of images containing only English characters, accuracy was high in the implemented method. However, when we applied our method to an image where Korean and English were mixed, we could see a drop off in performance. Through this result we can confirm that a different detection engine is required for the detection of multiple languages.



Figure 6. The test data set and experimental results

Table 1. Implemental and detected results

| Test data | Detected Text Information |
|--------------|---------------------------|
| Figure 6-(a) | Cyclists AHEAD SLOW |
| Figure 6-(c) | AHEAD |
| Figure 6-(e) | PED XING |
| Figure 6-(g) | 천천히 SLOW |

4.2. Phase II

TTS is a type of speech synthesis program that makes the contents of computer documents, such as help files and web pages, sound human-readable. TTS can also read image information for people with visual impairments. There are many TTS products on the market, including Read Please 2000, Proverbe Speech Unit, and TextAloud. Lucent and AT & T have their own products called “Text-to-Speech.” In this research, we applied Microsoft Speech API’s speech to text functionality in consideration of efficient aspects [26]. Figure 7 shows simplified block diagram for TTS function. This system works well for our test data.

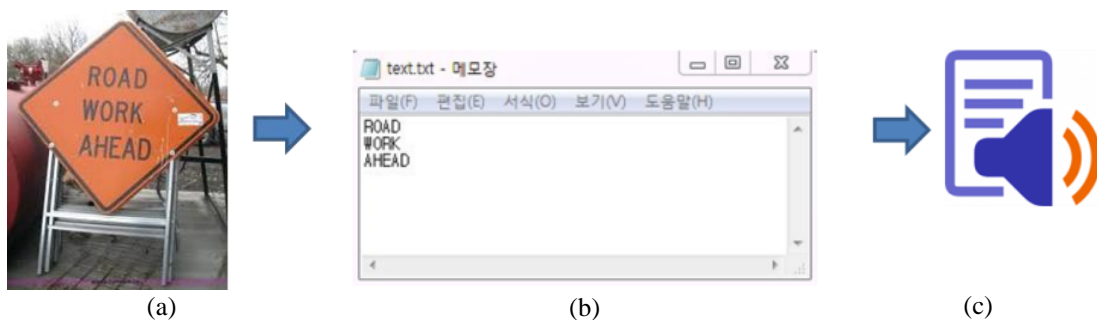


Figure 7. The block diagram for text-to-speech process. (a) input image, (b) captured letters on notepad, and (c) speech out with text.txt file

5. CONCLUSION

In this paper, we proposed a character recognition system for visual impaired persons and added in a voice guidance application implementation process. Our goal is to add a character recognition system for the visually impaired to the smartphone terminal through H/W production. We proposed a method to detect the characters and apply the voice guidance from the images captured in Phase I. Using the features of MSER and SWT, it showed the possibility of extracting the letters of road signs, and in Phase II, it showed the result of delivering the extracted text information to the users.

However, in this paper, we can look at various future points to consider. So far, the road guidance service application for the visually impaired has been developed using a blue pitcher communication based H/W auxiliary device such as a beacon in a specific area. With the development of IoT technology, the objects and data related to IoT have been increased greatly. It is necessary to develop a camera-based video information guidance service app that utilizes a voice guidance service technology that provides appropriate information. In addition, artificial intelligence techniques such as image recognition, natural language processing, and natural language generation can be used to enable blind people to live a more comprehensive and productive life.

ACKNOWLEDGEMENTS

This research was supported by the Daegu University Research Grant, 2017.

REFERENCES

- [1] A. A. Panchal, et al., "Character detection and recognition system for visually impaired people," *Proceedings of IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology*, India, pp. 1492-1496, 2016.
- [2] Abdurrasyid, et al., "Detection of immovable objects on visually impaired people walking aids," *TELKOMNIKA Telecommunication Computing Electronics and Control*, vol. 17, pp. 580-585, 2019.
- [3] D. A. Mohammed, et al., "Off-line handwritten character recognition using an integrated DBSCAN-ANN scheme," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 14, pp. 1443-1451, 2019.
- [4] O. O. Oladayo, "Yoruba Language and Numerals' Offline Interpreter Using Morphological and Template Matching," *IAES International Journal of Artificial Intelligence*, vol. 3, pp. 64-72, 2014.
- [5] I. F. Bt. Hairuman and O. M. Foong, "OCR signage recognition with skew & slant correction for visually impaired people," *Proceedings of IEEE International Conference on Hybrid Intelligent Systems*, Melacca, Malaysia, pp. 306-310, 2011.
- [6] M. E. Pollack, "Intelligent Technology for an Aging Population: The Use of AI to Assist Elders with Cognitive Impairment," *AI Magazine, American Association for Artificial Intelligence*, vol. 26, pp. 9-21, 2005.
- [7] J. Matas, et al., "Robust Wide Baseline Stereo from Maximally Stable Extremal Regions," *Proceedings of the British Machine Vision Conference, BMVC*, pp. 384-393, 2002.
- [8] Forssen and Per-Erik, "Maximally Stable Colour Regions for Recognition and Matching," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, CVPR, Minneapolis, USA*, 2007.
- [9] Q. Ye and D. Doermann, "Text Detection and Recognition in Imagery: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, pp. 1480-1500, 2015.
- [10] T. H. Rim, et al., "Influence of visual acuity on suicidal ideation, suicide attempts and depression in South Korea," *British Journal of Ophthalmology*, vol. 99, pp. 1112-1119, 2015.
- [11] Exploring Markets for Assistive Technologies for the Elderly, UN Gendered Innovations. Available: <https://genderedinnovations.stanford.edu/case-studies/robots.html>.
- [12] Global Elderly and Disabled Assistive Devices Market, "Coherent Market Insights," 2017.
- [13] N. Otsu, "A Threshold Selection Method from Gray Level Histogram," *IEEE Transactions on System*, vol. 19, pp. 62-66, 1979.
- [14] G. Park, et al., "A Study on Enhanced Binarization Method by Using Intensity Information," *Proceedings of the Spring Conference of the Korea Multimedia Society*, pp. 441-445, 2003.
- [15] K. Munadi, et al., "Improved Thresholding Method for Enhancing Jawi Binarization Performance," *Proceedings of 14th IAPR International Conference on Document Analysis and Recognition, Kyoto, Japan*, vol. 1, pp. 1108-1113, 2017.
- [16] R. Smith, "An Overview of the Tesseract OCR Engine," *Proceedings of 9th International Conference on Document Analysis and Recognition, Brazil*, 2007.
- [17] A. Abdulkader and M. R. Cassey, "Low cost correction of OCR errors using learning in a multi-engine environment," *Proceedings of IEEE International Conference on Document Analysis and Recognition, ICDAR2009*, pp. 576-580, 2009.
- [18] T. M. Breuel, et al., "High-performance OCR for printed English and fraktur using LSTM networks," *Proceedings of IEEE International Conference on Document Analysis and Recognition, ICDAR2013, Washington, USA*, 2013.
- [19] M. Heliński, M. Kmiecik, T Parkoła, "Report on the comparison of Tesseract and ABBYY FineReader OCR engines," *Impact*, 2012.

- [20] R. Shilkrot, J. Huber, W. M. Ee, P. Maes, S. C. Nanayakkara, "FingerReader: A Wearable Device to Explore Printed Text on the Go," *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI 2015, Seoul, Korea, 2015.
- [21] M. Waisbourd, O. M. Ahmed, J. Newman, M. Sahu, D. Robinson, L. Siam, C. B. Reamer, T. Zhan, M. Goldstein, S. Kurtz, M. R. Moster, L. A. Hark, L. J. Katz, "The Effect of an Innovative Vision Simulator (OrCam) on Quality of Life in Patients with Glaucoma," *Journal of Visual Impairment & Blindness*, Vo. 113, No. 4, pp. 332-340, 2019.
- [22] Korea Employment Development Institute, "2017 Disabled Statistics (Korean)," 2017.
- [23] WHO, "Global data on visual impairments," 2010.
- [24] L. Li and C. L. Tan, "Character Recognition under Severe Perspective Distortion," *Proceedings of IEEE 19th International Conference on Pattern Recognition, ICPR2008. Tampa, USA*, 2008.
- [25] K. Wang, et al., "End-to-end scene text recognition," *Proceedings of International Conference on Computer Vision, ICCV2011, Barcelona, Spain*, 2011.
- [26] Z. Zhang, "Microsoft Kinect sensor and its effect," *IEEE MultiMedia*, Vol. 19, No. 2, pp. 4-10, 2012.

BIOGRAPHIES OF AUTHORS



Jaejoon Kim received the MS and Ph.D. degree from the Department of Electrical Engineering, Iowa State University, USA. He received Bachelors from Department Electronics Engineering and Mathematics, Hanyang University, Korea. From 2001 to 2002, he was a senior researcher at ETRI (Electronics and Telecommunication Research Institute). He is currently a professor at Daegu University, Republic of Korea. His research interests include image processing, neural networks and non-destructive evaluation.